

## **CMA project update and response to Spend Network recommendations**

### **Background**

Working with contractors Spend Network, the CMA has developed a tool to help public procurers identify suspicious bid behaviour that might indicate the presence of a cartel. The tool consists of algorithms against which procurement data, specifically tender documentation and bid prices, can be tested; a positive result on one or more test could be cause for further investigation of that tender.

As part of this work Spend Network produced a report (February 2016, see below) setting out their work in identifying indicators associated with bid-rigging, selection of algorithms and the development of the tool. The report also summarised some of the challenges Spend Network encountered during the project and made some recommendations to the CMA for the next stages of the project.

This brief note sets out the CMA's activity on this project since receiving the report and responds specifically to Spend Network's recommendations.

We would like to acknowledge, with huge gratitude, the time and effort that procurement teams have devoted to helping us develop this tool.<sup>1</sup>

Project development since February 2016

Before considering the recommendations, it is worth considering how the project has developed since the report was delivered.

### ***A freely available tool***

The CMA has made the tool freely available to procuring authorities as a downloadable app to use within their own systems. This model (as opposed to a CMA-hosted tool) reflects the UK's distributed procurement arrangements. Putting the tool directly in the hands of the procurers allows the tool to take on a life of its own; authorities and e-procurement platforms will be able to absorb this sort of data analysis process into their internal systems rather than having to use a separate, centralised system. This, in turn, increases the scope for future tailoring and development of the tool.

This distribution model should help to build awareness of cartel risks across procurement officials but still result in intelligence sharing and enforcement leads for the CMA.

Crucially, this also means that procurers do not need to share their data for analysis. The reluctance of authorities to share data was a key obstacle to the early stage development of the tool.

---

<sup>1</sup> Thank you to all the organisations which have supported and helped us with this project, with particular thanks to the Cabinet Office, Cheshire West, Cornwall, Department of Health, Department of Transport, Devon, Durham, East Surrey and Sussex, Government Digital Service, Halton, Lambeth, NHS Scotland, Northern Ireland, Perth and Kinross (Scotland), Warrington, and the Welsh Government.

This distribution model carries the risk that, by making the workings of the tool visible, the greater understanding of cartel detection techniques might be abused. But we think the potential benefits of the tool being taken on, development and promoted outweigh this risk.

### **Front end**

Spend Network successfully delivered a set of algorithms, but given the decision to pursue a distributed model, the CMA decided to invest in developing a straightforward, easy to use front end for the tool. The priorities were to minimise the amount of data entry needed, make the interface user friendly and ensure the results were easy to interpret. Following a competitive tender process, the CMA contracted Symfanet to deliver the front end.

## **Spend Network Recommendations**

### **Summary table**

Spend Network recommendation	CMA response
1: Continue testing and calibrating the tool using more data	Complete. The CMA continued to work with Spend Network to secure and analyse further contracts. As a result of this work, users can now adjust thresholds and weightings of individual tests within the tool to best reflect their specific circumstances.
2. Resolve legal issues	The tool has developed in a way that obviates the legal issues identified as procuring authorities do not have to share their data in order to use the tool.
3. Create a platform to collect and process the data	Reflecting the situation in the UK, the tool has developed in a way that does not require the CMA to have an ongoing co-ordinating role. However, should others want to share experience and/or create such a central platform we would encourage that.  However, as recommended by Spend Network, the CMA has supported the introduction of the Open Contracting Data Standard by Government in order to make it easier to compare and process files in the future.
4. Collaborate with other interested parties	The CMA has collaborated with central and local government, international competition agencies and others. We have used their feedback to help shape the development of the tool, and will continue to value user feedback, including where they have adapted the algorithms and tool to suit their circumstances.  We have not, as yet, taken Spend Network up on their offer to host a workshop. We will consider the need for a workshop to review distribution, use and development of the tool later in 2017.

## **Next steps**

Public procurement and the detection and deterrence of cartels is an ongoing priority for the CMA. We will continue to promote the tool and the various supporting materials.

Although the CMA is not currently planning to do any further work on analysing data tested by the tool, or in further developing the tool itself, we do not want to preclude doing so. We will review use of the tool and any significant developments in early 2018. We will facilitate sharing of experience and insights between users, Spend Network and other interested parties; and may consider how the tool and this data analysis approach might best be developed in future.

# **Report on bid-rigging algorithms**

**The use of public sector procurement  
information to identify bid-rigging in public  
procurement markets.**

Prepared for the Competition and Markets Authority by Spend Network  
February 2016

**Spend Network** | 

## Table of Contents

- [Table of Contents](#)
- [Introduction to the project](#)
  - [Introduction](#)
  - [Scope](#)
  - [Project Stages](#)
  - [Publication](#)
  - [Findings](#)
- [Partner participation](#)
  - [Introduction](#)
  - [Overview of partner involvement](#)
  - [Partner participation](#)
- [Literature review](#)
  - [Introduction](#)
  - [The Literature](#)
  - [Trends in Literature](#)
  - [Identifiers of cartel activity](#)
  - [Limitations](#)
  - [Outcome](#)
- [Selecting algorithms](#)
  - [Introduction](#)
  - [Options presented at Workshop 1](#)
- [Gathering Data](#)
  - [Introduction](#)
  - [Legal issues](#)
    - [CMA legal and data security issues](#)
    - [Partner legal issues](#)
    - [Summary of partner data](#)
  - [Data transfer methods](#)
  - [Conclusion](#)
- [Processing data](#)
  - [Introduction](#)
  - [Processing the Data](#)
  - [Issues](#)
    - [Data Formats](#)
    - [Manual effort](#)
    - [Council-level procurement processes](#)
  - [Conclusion](#)
- [Initial algorithm results](#)
  - [Introduction](#)
  - [Algorithms as presented to Workshop 2](#)
  - [Attendee feedback](#)
  - [Outcomes of Workshop 2](#)
- [Amending the algorithms](#)
  - [Introduction](#)

[Sample tests](#)

[Further development](#)

[The Tool in Practice](#)

[Introduction](#)

[Inputs](#)

[Data](#)

[Test weighting](#)

[The algorithms in practice](#)

[Metric Scores](#)

[Tests applied](#)

[Outputs](#)

[Further Amendments](#)

[Using the tool](#)

[Introduction](#)

[Recommendations](#)

[Introduction](#)

[Recommendation 1: Continue testing and calibrating the tool using more data](#)

[Recommendation 2: Resolve legal issues](#)

[Recommendation 3: Create a platform to collect and process the data](#)

[Recommendation 4: Collaborate with other interested parties](#)

[Concluding Remarks](#)

# Introduction to the project

## Introduction

This document presents Spend Network's work to use algorithms to identify identifiers that could be associated with bid-rigging for the CMA. It details the reasoning for selecting the algorithms and metrics used in the final tool and introduces this tool. It also summarises some of the challenges encountered during the project and makes recommendations for the future.

## Scope

Spend Network were contracted to:

...‘design and test a methodology which can be applied to information gathered by public sector procurement organisations in order to identify potential anti-competitive behaviour amongst public sector suppliers.’

The aim of the project was to attempt to prove the concept that more sophisticated analysis of government procurement data could be used to identify potential instances of bid-rigging. This would involve creating and testing analytical approaches that could be applied to public sector procurement data to identify potential bid-rigging. A tool would be developed to identify warning signs for cartelised behaviour.

It was recognised that testing and developing a usable methodology would be an iterative process, but the work was focused on proving the concept. Therefore, any further investigation of collusion would be out of scope.

In the first meetings of the project, it was agreed that the project would focus on tender and bid submission data from partner local government organisations and that the market profiled for this project would be construction.

## Project Stages

The following stages and milestones were agreed at the start of the project:

- Kick off meeting
- Engage partners
- Literature review
- Prioritisation analysis and development of a long list of algorithms
- Workshop 1: Methodology kick off
- Data gathering and processing
- Data linking and augmentation
- Develop algorithms
- Workshop 2: Refinements
- Amends
- Workshop 3: Project findings and recommendations
- Project reporting

## Findings

Spend Network were able to successfully use algorithms to analyse data from government procurement information, specifically tender documents and bid submissions. However, the tool must be used on a larger dataset before we can fully evaluate its effectiveness. We believe that, with a large dataset, it is highly likely that this could be used to identify suspicious activity in public tenders that could be associated with cartels.

The tool is a set of algorithmic tests that can be used on tender and bid data to measure events to create indicators of suspicious behaviour and alert an investigating authority. The investigating authority can adjust the thresholds of the test metrics and the weightings of the tests to adapt it to different markets. Further tests for suspicious behaviour can be added as required.

More data is necessary to calibrate the tool, including adjusting it for different categories like region or industry. This further testing should include data from known cartels to test the tool's ability to identify suspicious documents.



# Partner participation

## Introduction

The CMA contacted councils and other local government bodies, inviting them to join this project, and passed on the details of the entities expressing an interest to Spend Network. These entities became our project partners and many supplied data to the project and feedback on the final tool.

The partners indicated they were interested in supplying the project with tender and bid data. Not all the initial partners were able to take part in the final project. We progressed using data from eight English local authorities.

## Partner participation

As well as supplying data, the partners were also able to help us by providing practical expertise on how procurement processes work in practice. They were also able to make suggestions based on previous experience in tracing cartel activity.

One local authority was able to co-ordinate activity with other partners, an approach that could be used in the future to expand the selection of data used in this research. More information on Spend Network and the CMA's work with the partners, particularly concerning legal issues, can be found in the section [Partner legal issues](#).

# Literature review

## Introduction

Before beginning the investigation, Spend Network engaged in a review of the literature on cartel activity to identify common trends in spotting cartels. These trends could be used to develop algorithms that could be used as identifiers of potential cartel activity.

The initial contract stated:

‘The methodology will be based on insight from the economic literature on the circumstances in which bid-rigging is more likely to arise (such as a small number of firms, simple products etc) and evidence collected from historic bid-rigging cases (such as a bids coming from the same IP address, common mistakes in bids etc).’

## The Literature

The CMA provided a reading list that included:

- [Australian Competition and Consumer Commission: Cartels Deterrence and detection a guide for government procurement officers](#)
- [Fair Trading Commission Barbados: detecting, mitigating and fighting bid-rigging in public procurement](#)
- [Canadian Competition Bureau: bid-rigging awareness and prevention](#)
- [Competition Commission of India: bid-rigging](#)
- [The Competition Authority Ireland: the detection and prevention of collusive tendering](#)
- [Korean Fair Trade Commission: ICN presentation on measure to detect cartels other than leniency](#)
- [OECD: guidelines for fighting bid-rigging in public procurement](#)
- [South African Competition Commission: how to detect and report collusive tendering](#)
- [Swedish Competition Authority: twelve ways to detect bid-rigging cartels](#)
- [OECD - Fighting Hardcore cartels](#)
- [http://www.mihalyfazekas.eu/Fazekas-Toth-King\\_Composite%20indicator\\_2013.pdf](http://www.mihalyfazekas.eu/Fazekas-Toth-King_Composite%20indicator_2013.pdf)

## Trends in Literature

Spend Network identified two primary indicators of bid rigging from the literature:

1. The market has to be ‘riggable’ in that the cartel has to be confident that they are able to secure an agreement with all of the viable bidders. Competitive markets, such as web design, are hard to rig as there are so many firms ready to bid.
2. The rigged contract has to generate higher than average revenues for the suppliers, both as a reward for risk and to ensure that cartel continues to function.

## Identifiers of cartel activity

Based on these findings, Spend Network honed in on these identifiers of cartel activity:

- A limited market with few suppliers:<sup>1</sup>
  - Cartels with few members are more efficient, as they are easier to maintain, carry less risk of detection and also require fewer payments between cartel members, so they can secure additional profit without deploying an egregious pricing regime.
- Prices notably higher than the market value, including prices from the same suppliers over a number of tenders and prices in specific tenders:<sup>2</sup>
  - The spread of prices and the pricing of the winning tender could all imply collusion, particularly if all these prices were higher than the market rate.
  - Line item costs are manipulated to raise the prices above the typical market value.
- Uncompetitive markets in which cartels can thrive:<sup>3</sup>
  - For a cartel to operate, there has to be a certain level of lock out, with a small, manageable group of suppliers confident that they can deliver on a contract. Therefore we should look to establish an understanding of which categories suffer from a limited market, such as frameworks.
  - Tenders with a lower number of response than are typically received for similar tenders across the market and patterns with the same suppliers repeatedly bidding for tenders with the same buyer, would indicate an uncompetitive market in which a cartel could thrive.
- Variations in bid quality as cartel participants decide which bid should win before the competition takes place:<sup>4</sup>
  - The bid that is chosen to win and is made appealing. It is well-written and the price, while still high in comparison to the general market, is still considerably lower than the other bids for the tender.
  - Knowing that their bids are unlikely to win, the other suppliers do not make the effort to submit a high-quality bid.
- Signs of collaboration between bidders, including indications that the bids were written with knowledge of the other suppliers' submissions:<sup>5</sup>

<sup>1</sup> Australian Consumer and Competition Authority 2009, *Cartels Deterrence and Detection: a guide for government procurement officers*. Available from <<https://www.accc.gov.au/system/files/Cartels%20deterrence%20and%20detection%20-%20and%20checklist.pdf>> [4 January 2016]; OECD 2009, *Guidelines for fighting bid-rigging in public procurement*, Available from <<http://www.oecd.org/competition/cartels/42851044.pdf>> [4 January 2016]

<sup>2</sup> The Competition Authority Ireland 2009, *The detection and prevention of collusive tendering*. Available from <<http://www.tca.ie/images/uploaded/documents/Booklet%20-%20The%20Detection%20and%20Prevention%20of%20Collusive%20Tendering.pdf>> [4 January 2016]; Konkursverket, Swedish Competition Authority, Honesty Pays! How to detect and give tip-offs about bid-rigging cartels in public procurement. Available from <<http://www.konkurrensverket.se/globalassets/english/publications-and-decisions/honesty-pays-how-to-detect-and-give-tip-offs-about-bid-rigging-cartels-in-public-procurement.pdf>> [4 January 2016]

<sup>3</sup> Fair Trading Commission Barbados 2011, *Detecting, mitigating and fighting bid-rigging in public procurement*. Available from <[http://www.ftc.gov.bb/library/2011-02-07\\_ftc\\_guidelines\\_checklist\\_procurement.pdf](http://www.ftc.gov.bb/library/2011-02-07_ftc_guidelines_checklist_procurement.pdf)> [3 January 2016]; Australian Consumer and Competition Authority 2009, *Cartels Deterrence and Detection: a guide for government procurement officers*. Available from <<https://www.accc.gov.au/system/files/Cartels%20deterrence%20and%20detection%20-%20and%20checklist.pdf>> [4 January 2016]; OECD 2009, *Guidelines for fighting bid-rigging in public procurement*, Available from <<http://www.oecd.org/competition/cartels/42851044.pdf>> [4 January 2016]

<sup>4</sup> Konkursverket, Swedish Competition Authority, Honesty Pays! How to detect and give tip-offs about bid-rigging cartels in public procurement. Available from <<http://www.konkurrensverket.se/globalassets/english/publications-and-decisions/honesty-pays-how-to-detect-and-give-tip-offs-about-bid-rigging-cartels-in-public-procurement.pdf>> [4 January 2016]; OECD 2009, *Guidelines for fighting bid-rigging in public procurement*, Available from <<http://www.oecd.org/competition/cartels/42851044.pdf>> [4 January 2016]

<sup>5</sup> Australian Consumer and Competition Authority 2009, *Cartels Deterrence and Detection: a guide for government procurement officers*. Available from <<https://www.accc.gov.au/system/files/Cartels%20deterrence%20and%20detection%20-%20and%20checklist.pdf>> [4 January 2016]; Konkursverket, Swedish Competition Authority, Honesty Pays! How to detect and give tip-offs about bid-rigging cartels in public

- Text similarity between bids could imply that the same person was responsible for more than one bid or that suppliers copied from each other's bids.

## Limitations

The literature revealed that most indicators are inconclusive, they could be evidence of cartel activity but there could also be an innocent explanation.

- A low number of bidders could be a sign of an uncompetitive cartelised market, for example, it would be normal for very few bidders to respond to aerospace tenders. Low numbers of bidders could also be the result of a restrictive specification making bidding unappealing for all but a handful of very specialised suppliers.
- Frameworks make it difficult to spot cartelisation as the number of suppliers is already low and it is difficult to get oversight of mini-competitions.
- While this low number of bidders can reduce the cost of the procurement process, this benefit can prove illusory as the lack of competition can ultimately drive prices higher.

Some cartels may operate on an opportunistic basis, where suppliers simply agree to submit an inflated bid on a single tender, with an agreement that the winning bidder will make a facilitation payment to losing bidders. These more casual cartels are hard to identify.

## Outcome

The identifiers of cartel activity listed above were used to devise a long list of algorithms that could possibly allow the person applying the algorithms on tender or bid data to spot these identifiers. Thresholds were set on these tests to avoid false positives. The algorithms can be used on individual tenders to identify signs that could indicate more casual collusion.

---

procurement. Available from <<http://www.konkurrensverket.se/globalassets/english/publications-and-decisions/honesty-pays-how-to-detect-and-give-tip-offs-about-bid-rigging-cartels-in-public-procurement.pdf>> [4 January 2016]; OECD 2009, *Guidelines for fighting bid-rigging in public procurement*, Available from <<http://www.oecd.org/competition/cartels/42851044.pdf>> [4 January 2016]

# Selecting algorithms

## Introduction

Based on the literature review, Spend Network devised algorithmic tests that would check for these indicators of cartel behaviour:

- An uncompetitive market
- Inflated prices
- Collusion between buyers
- Varying effort in bid submissions.

The resulting options were presented to a workshop of experts from the CMA and other authorities investigating cartelisation. They discussed the options and the CMA and Spend Network used this feedback to select the final algorithms. This section outlines all the options presented at the workshop, which ones were selected, which ones were rejected and the reasons for these decisions.

## Options presented at Workshop 1

This table outlines the options presented at the first workshop and why they were accepted or rejected. It was agreed at the CMA follow-up meeting to choose a selection of indicators and document based indicators to ensure a variety of tests.

Algorithm	Description	Accepted	Reason
Low number of bidders	A test that assesses whether the number of bidders was below average for the category.	Y	Easy to test and a good indicator of an uncompetitive market.
One bidder / Tender boycott	A test that monitors tenders that receive just one bid.	Y	Easy to test and a good indicator of an uncompetitive market.
Market competitiveness	A test to determine whether a tender or group of tenders are routinely operating at below the market average in terms of tender responses by comparing the tenders of one buyer to the wider market.	Y	Understanding the market is a vital to knowing whether a cartel could function in a given area. Comparing the tenders from one buyer to the wider market can reveal a low number of suppliers and high prices, key indicators of cartelisation.

Over pricing	A test to determine whether a tender is overpriced, using the average spend data for the chosen supplier or category.	Y	High prices are a significant indication of cartel activity.
Outlier pricing in a tender	A test for pricing where a small number of bidders have 'broken ranks' and submitted a normal market price against cartel members who have inflated their prices.	Y	Highlights how far prices have varied from the market. Requires only the data from each tender and adds further intelligence to pricing data.
Mathematical similarity in costing	Analysis of line item costs in tenders to see if the line item costs in bids for the same tender are mathematically similar.	Y	Indicates that the suppliers manipulated the pricing in a way that fails Benford's law (see details here: <a href="http://mathworld.wolfram.com/BenfordsLaw.html">http://mathworld.wolfram.com/BenfordsLaw.html</a> ). Useful to test where line item pricing is available.
Document authorship metadata (Word document)	A test to determine if a document was authored by someone other than the supplier making the submission.	Y	Useful check on the provenance of the documents. Easy to access the data from Word documents.
Low endeavour submission (Word document)	A test using the metadata stored in Microsoft Word documents that details the changes and submissions made to a document, as a means to indicate the amount of effort that went into compiling a tender prior to submission.	Y	Indicates collusion as suppliers have made a prior agreement on which bid will be successful and so put little effort into the other bids. Easy to test where Word documents exist.
Similar prices	A test to evaluate the similarity in prices across different tenders	Y	Shows if a cartel has agreed on the price of the bid it planned to win and made it the clear choice by lowering the price in comparison to the other submissions from cartel members, whose prices would all be similar. Requires only the data from each tender and adds further intelligence to pricing data.

Submission similarity	A test to determine whether submissions by different suppliers are suspiciously similar	Y	High text similarity would imply collusion in creating bids and therefore would be a very strong indicator of cartelisation.
Document authorship metadata - PDF	A test to determine if a document was authored by someone other than the supplier making the submission.	N	Metadata rarely exists for PDF documents.
Submission timing	A test to establish whether there is any unlikely pairing in the timing of bids that could be suspicious. If the timing of bid returns are clustered much more tightly than average it may indicate that suppliers are conversing before submitting their bids, or that bids were submitted by the same individual.	N	It was feared that the inevitably higher volume of bid-submissions near deadlines could distort this analysis. This also relied on the data on submission timings being stored either by procurement portals or by the bid recipient.
Low endeavour submission (online questionnaire)	A test to count the number of log entries in a tender and to see if all the suppliers but one have made little or no effort when submitting their bids.	N	This required bids to be submitted via online questions, a method not used by any of our partners.
Suspicious pattern of awards: Repeated awards to incumbents	Geographic, sector and other pattern analysis to see if there is a pattern of incumbents retaining the business in a market.	N	To make a valid analysis, this would have required data from more partners than we had and over a longer duration of time. It is very difficult to determine the geographic area over which a cartel operates we cannot define the set area of cartel activity. Data for the whole country is required before these geographic distinctions would be possible.
Suspicious pattern of awards: Suppliers take turns	Geographic, sector and other pattern analysis to see if there is a cycle of a small group of suppliers winning the business in a market.	N	To make a valid analysis, this would have required data from more partners than we had and over a longer duration of time. It is very difficult to determine the geographic area over which a cartel operates we cannot define the set area of cartel activity. Data for the whole country is required before these geographic

			distinctions would be possible.
Suspicious pattern of awards: One supplier dominates	Geographic, sector and other pattern analysis to see if a single supplier is winning all of the business in a market.	N	To make a valid analysis, this would have required data from more partners than we had and over a longer duration of time. It is very difficult to determine the geographic area over which a cartel operates we cannot define the set area of cartel activity. Data for the whole country is required before these geographic distinctions would be possible.
Typo matching in bids	Analysis of typographic errors in the texts of bid submissions to see if the same mistake is repeated in bids from different suppliers to the same tender. This would indicate the same author for all submissions.	N	A dictionary and any other relevant acronyms and names would need to be added to the text analysis tool. While this is a valid analysis, the text similarity algorithm was deemed preferable as it already identifies similar words and phrases in the documents but has a wider scope than just typographic errors.
Submission source	Test to see if tenders from rival suppliers were submitted from the same IP.	N	It would have been difficult to gather this data and the algorithm only identifies a very specific mistake by cartelists.
Login source	Where submissions are made via webforms, it may be possible for the same supplier to be submitting data on behalf of other suppliers. Monitoring IP addresses will test for this.	N	It would have been difficult to gather this data and the algorithm only identifies a very specific mistake by cartelists.
Submission authorship tests	Use of pre-existing algorithms to test language patterns that indicate texts are written by the same author.	N	The text similarity algorithm already identifies similar words and phrases in the documents.
Losing bid quality analysis	A test that analyses patterns in the submissions of a single supplier to determine whether or not they are deliberately submitting poor quality bids at different times.	N	This depends on data for the same supplier across multiple tenders and Spend Network could not guarantee access to this sort of data. Bid quality is already being tested using metadata.
Tender	A test that predicts the	N	Tender specifications can vary for a wide



competitiveness	number of responses based on the nature of the tender specification.		variety of reasons depending on what is being procured. Accounting for this sort of variation makes it difficult to set benchmarks for quality, therefore making it difficult to run this test.
Losing price analysis	A test that monitors pricing data to see if suppliers are bidding to lose by comparing the pricing variations between bids.	N	Depends on data for the same supplier across multiple tenders. Pricing data is already being analysed by various other algorithms.
Pricing immunity	Prices in certain area remaining stable or increase, while prices elsewhere are falling. Testing this requires pattern analysis to check for a distortion of prices in a specific market.	N	This would require data over a longer period of time than we had available. Pricing and markets were covered by other tests.

# Gathering Data

## Introduction

Spend Network gathered data from six partners in August and September, through secure channels.

This section also highlights the issues encountered in this stage of the project. There were several weeks of delay due to resolving legal and data security issues. Delays due to absence were a common feature in much of Spend Network’s data collection activities as this work took place in August. This slowed the progress of the project. In many cases, councils were not able to begin gathering and transferring data until September.

## Legal issues

### CMA legal and data security issues

The CMA and Spend Network needed to find a way to share data while protecting the interest of the local authority procurement teams in keeping some commercial information confidential. Agreeing to anonymise data before storing it, and providing assurances that the CMA considered the data to be commercial in confidence was sufficient for some partners. However, some potential partners felt unable to supply data unless they received a more robust guarantee.

It was agreed that data could be transferred to Spend Network via a secure Cloud from which it would be downloaded then deleted. The data would be thoroughly anonymised before being analysed and the CMA would not take ownership of the data.

## **Partner legal issues**

The CMA's data security team approved the data transfer process at the end of July. However, in most cases, the partners required that the data transfer method be approved by their own IT security and legal teams. One local authority was unable to get the project approved by its own IT security team so we were unable to access this data.

## **Data transfer methods**

All the councils used secure methods to transfer data to Spend Network. There was a range of methods used for this purpose. The method of data transfer varied depending on the way councils stored data and the approach the council required for secure data transfer. Some council procurement departments were blocked from using our online tool by internal IT policies so they needed to agree new methods to transfer data. Some also had difficulty accessing and using the encryption tools recommended by Spend Network and the CMA for secure transfer of data. Co-ordinating this data collection was made even more difficult due to absences in with council procurement, legal and IT teams in August, so the deadline for the final transfer of data was set for late September. Therefore, the process of transferring data was time consuming both for the council and for Spend Network.

## **Conclusion**

Five councils provided data on 55 tenders, with a total of 237 documents. However, this was the lengthiest part of the project due to the time taken to find a legal resolution to Freedom of Information and data security issues.

# Processing data

## Introduction

Once the data was collected, Spend Network began to process it could be used for algorithmic testing. The data had to be made machine-readable and each bid or tender had to be in the same format. While this was a time-consuming process, it allowed Spend Network to become familiar with the data and how it could be synthesised for analysis.

## Processing the Data

A .csv file was created for each document, containing data on its format and content.

Heading	Description
Document type	Bid or Tender?
Date	Deadline of tender or submission date of bid
Buyer Code	Code created as an identifier by Spend Network
Supplier Code	Code created as an identifier by Spend Network
Tender ID	Tender ID created as an identifier by Spend Network
Text?	Is there plain text of the document available?
Price	Final total price
Winner?	Was this the winning bid?
Metadata?	Was there document metadata?
Time editing	How long was spent editing the document?
Number of edits	How many edits were made to the document?
Author Name	The author name in the metadata
Suspicious?	'Y' if the signatory is different to the author, 'N' if not
Ben	First digits of line items

An additional .txt file contained the anonymised plain text of the documents was also produced where possible. Scanned documents could not be converted into text.

The data was processed separately from the algorithm development, so the developers of the algorithms never saw the original tender and bid documents. The files were completely anonymised, so the developers were unaware of the real identities of the buyers or the suppliers. This prevented any chance of the results being prejudiced while also adding to the security of the data.

## Issues

There were numerous issues with processing the data and, while some of these issues were resolved during this stage of the project, others will need further attention if more data is gathered for further work.

### Data Formats

The format of the tender documents and bids made it difficult to process the data. Most submissions involved large volumes of PDF files that had to be converted into text. Spend Network was able to develop a script to convert the data, but document format variations did cause issues, especially with tabulated data. A script was also used to transform the .msg files submitted by two local authorities, saving attachments in their original form and translating email messages to text.

Unfortunately, other issues with the data could not be resolved so easily. Some suppliers had submitted their bids as scanned documents that meant that they could not be read by machine or converted to text. One local authority submitted files with the supplier names redacted. This made it impossible to tell if a supplier for one council was also featuring in their data, so we were unable to use this data.

### Manual effort

The manual effort was essential to find and process data on pricing. Each tender needed to be opened and reviewed manually to find the total price. When finding line item prices, in some cases there were a number of different tabs had to be opened from Bill of Quantities spreadsheets, and the data copied over.

### Council-level procurement processes

The way in which councils manage construction projects could also distort the analysis. Many of the tenders submitted by three local authorities were part of framework agreements. This artificially reduced the number of companies bidding for contracts. Moreover, two of these authorities conducted a pre-assessment evaluation and nearly always invited only three suppliers to bid for the contract. This may have affected our analysis of the typical number of bidders for a contract. Further difficulties with data from these authorities were caused by the new policy of offering small contracts on rotation after assessing a small number of companies in a framework agreement for suitability. This tendering process did not fit to our method of evaluation, which meant much of the data could not be used in the final analysis.

## Conclusion

By processing the data, Spend Network created a .csv file for every document submitted and a .txt file wherever possible. Processing the data was made complicated by the range of formats used in tender and bid documentation. A great deal of manual intervention was needed to extract document metadata and pricing data. Automating this process is possible but will require further development.

# Initial algorithm results

## Introduction

The results of the initial algorithmic tests were presented to the second workshop, which was attended by experts from the CMA and representatives of the partners. Spend Network presented the version of each algorithm and the challenges involved in developing them and using them on the council data.

Spend Network revealed the initial pass/fail scores for each tender as an anonymised table. The tenders that passed were ones which did not score highly for signs of potential cartel behaviour, while the failing tenders scored slightly higher for these signs of cartel behaviour and could potentially merit further investigation. For more information on scoring, see [The Tool in Practice](#).

## Algorithms as presented to Workshop 2

The algorithms were presented to the workshop with comments explaining how successful the development of these tests had been. Each algorithm was applied to the data gathered from the tenders and bids detailed above. However, sometimes not every bid or tender could be tried against every algorithm, for example, Benford's Law could only be tested when all bids for a tender had line item prices. A test was considered successful when the algorithms could be applied to give a score. When calibrated with more data, the scores from these tests could be used to detect indicators associated with potential cartel behaviour.

For some other tests there were concerns about the data, for example the sample size may have been overly limited or it may have required a significant amount of manual intervention to access the data. Furthermore, the market competitiveness test was not effective as we did not have a wide range of data across the whole market and so could not form a baseline. The table below shows the success of each algorithm:

Algorithm	Description	Comment
Low number of bidders	A lower than average number of bids	Successfully tested
One bidder (aka Tender boycott)	Only one bid in response to a tender	Successfully tested
Market competitiveness	Compares the pricing of the bid responses to the typical pricing in the market	Issues with forming a baseline for spending due to lack of data.
Outlier pricing in a tender	The winning price is significantly lower than the other bids	Successfully analysed, where sufficient data.
Mathematical similarity in costing	Use Benford's Law to check the line item costing of bids to find out when the numbers	Successfully analysed, limited

	<p>look made up. Benford's Law states that 1 will be the first digit in around 30% of numbers in a naturally occurring number set. If line item costs were manipulated by a cartel then they would not conform to Benford's law.</p>	dataset.
Document authorship metadata	By checking Word document metadata, find out if the author in the metadata is different from the person who signed the document.	Required manual analysis, data questionable.
Low endeavour submission	By checking Word document metadata, find out when a below average number of edits have been made or a below average amount of time has been spent preparing the document.	Successfully analysed, where data was available.
Similar prices	All the prices are comparatively high, with a clustering of the highest prices.	Successfully analysed, requires > 1 response
Text similarity	The texts of the bids are very similar.	Successfully analysed, would like to refine algorithms

These algorithms were presented at the workshop with comments on their application and usefulness. This enabled the attendees to evaluate the value of each test and to understand the further work needed to hone these algorithms.

In general, more data was needed to further test and calibrate the algorithms. The example of the market competitiveness algorithm demonstrates how important this is to create a viable tool. A picture of a typical market and its competitiveness could not be created with the current data. For each market, Spend Network would require many more tender documents and bid responses across many different regions over several years to build up an idea of how many responses would be expected in a normal, competitive market and the typical costs of similar tenders. Our current data analyses a small selection of tenders from a few buyers and so cannot be used to extrapolate a more general view of the market. Testing for market competitiveness using this data could result false positives as we cannot know if a market is truly uncompetitive without a comparison to known, healthy markets. The limited selection of tenders made it difficult to find instances in which like could be compared with like across different regions, making it impossible to determine whether costings were above or below typical market rates.

Testing the algorithms was restricted by other limitations in the data. Not all bids had line item costs and only Word documents contained document authorship and effort data. The lower than expected average number of bids for each tender weakened the distinction between the low number of bidders and one bidder (tender boycott) algorithms.

## Attendee feedback

The attendees gave feedback on the metrics and the algorithms. They provided examples from personal

experience of bid-rigging and how they perceived these algorithms could be used in a professional context. They also made suggestions of how to improve the tests and of metrics that could be added in future.

The attendees reviewed the algorithms and completed forms to show which ones they thought were the most significant indicators of cartel activity by giving numerical scores. A score of 1 indicates that the attendee viewed the algorithm as being very useful, while 5 showed that the algorithm was not considered useful at all. Where the option was left blank in the attendee's response, they were scored 6 as additional comments revealed that this indicated that the attendee had a low regard for the algorithm.

The scores and comments from the attendees revealed that Text Similarity was considered the best indicator of potential cartel activity, although it was acknowledged that it would probably only flag cartel activity in very rare circumstances. Low Endeavour Submissions and Mathematically Similar Pricing were also seen to be strong indicators of cartel activity.

Due to low sample numbers and other problems in the current data, there were some algorithms that the meeting attendees did not consider to be particularly useful now, while recognised that they had the potential to be more useful in the future with further development and/or more data.

- Market competitiveness was not seen as particularly useful as indicator right now, because problems linking spend and tenders meant that it was hard to determine competitiveness with the current data. However, if there was more data and it was linked, this metric could be very useful for determining where cartels could thrive.
- Document authorship metadata was not seen as a good indicator on its own, but the meeting attendees believed that linking it to social media, for example LinkedIn, could expose interesting connections between document authors.

## Outcomes of Workshop 2

Spend Network took the feedback from the attendees and used it to refine the algorithms to by weighting the ones considered most significant. The information gathered from the meeting also informed the setting of pass/fail thresholds for each metric and the creation of tests with multiple algorithms. This would all be used to present a methodology and tool to the final workshop.

# Amending the algorithms

## Introduction

Referencing the feedback from the Workshop 2, Spend Network refined the algorithms and developed new tests. Each of these tests were set a threshold for failure; weighted, so some would be a stronger indication of cartel behaviour than others, and scored, so a high score would require further investigation as it could be a sign of cartel behaviour.

## Sample tests

Some of the tests were simply triggered by crossing a threshold on individual metrics. For example, if the tender responses were to have very similar texts, this would result in a high score and the tender would be flagged as suspicious.

Other tests were developed using a combination of metrics. Spend Network drew up scenarios to describe cartel behaviour and show the combinations of metrics that would occur in these circumstances, for example:

- Market and pricing indicators:
  - A cartel operates in an uncompetitive market with few suppliers. All prices submitted by the cartel in response to a tender are above the market average. The cartel chooses one bid to win. If this bid is clearly lower in price than the other bids then the combination of these metrics is a powerful indicator of potential bid rigging.
  - Metrics:
    - Low number of bidders
    - Market competitiveness
    - Similar pricing
    - Outlier pricing
- Procedural indicators:
  - Knowing that their bids are likely to be rejected, the other suppliers will put very little effort into their bids, making few edits and spending little time on the documents. The line item prices will be set to ensure that all the bids are high and that one bid is lower than the others. These numbers may well be artificially manipulated and will fail the test for a naturally occurring dataset, so they will not conform to Benford's Law.
  - Metrics:
    - Low effort bids
    - Mathematically similar pricing
- Price manipulation by cartel:
  - The winning bid should be an obvious choice, so the price will be significantly lower than the other bids. The line item prices will be set to ensure that all the bids are high and that one bid is lower than the others. These numbers will be artificially manipulated and will not reflect normal market rates so they will not conform to Benford's Law.
  - Metrics:
    - Outlier pricing
    - Mathematically similar pricing
- The combination of the metrics mentioned in the above tests would score highly as a test and would be flagged for further investigation. The individual metrics would each flag on their own, but this test is to highlight the combination as particularly suspicious.

## Further development

Spend Network ensured that the final tests are configurable so that they can be developed for the future. More metrics can be added if they are found to be good indicators of cartel activity, particularly as a result of further testing with data from known cartels (see [Recommendation 1: Continue testing and calibrating the tool using more data](#) for more information). These new tests would need to be coded, but the structure of the code means once a new function is written and added to a file then it would become available to test.

More tests can be developed from different combinations of metrics as new cartel activities are observed. Thresholds can be altered as more data gives us a stronger indication of how non-cartel and cartel behaviour



affects the metrics. The weight given to each test can be changed if it is found to be a particularly strong or weak indicator of cartel activity.

# The Tool in Practice

## Introduction

This section aims to give an overview of how the tool works in practice. It will demonstrate how the data from each bid and tender document is used to create metrics for each tender. Metric scores are then applied and these are tested. The tests are weighted and final pass/fail results and suspicion scores are given to each tender. This output reveals which tenders may need further investigation.

Note that the examples in the charts shown in this section are for illustrative purposes only.

## Inputs

### Data

Data on each bid and tender document in csv form is processed to create overall data for each tender. This chart shows metrics for two sample tenders:

- Tender 1 had 5 bid responses, a mean price of 10,000, a standard deviation of 1000, a mean word count of 300, and a Benford's Law fit of 20%.
- Tender 2 had only one bid response, a mean price of 250,000, a standard deviation of 50,000, a mean word count of 5,000, and a Benford's Law fit of 90%.

Tender Data					
	# Bids	Mean Price	Standard Deviation	Mean Word Count	Benford's Law fit.
tender1	5	10000	1000	300	20 %
tender2	1	250000	50000	5000	90 %

### Test weighting

The user can set and adjust weightings for each test to make it have a stronger or weaker impact on the final results. For example, if Benford's Law was considered to be a very strong indicator of potential cartel activity, the weight applied to this test could be increased. This chart shows a simplified example of test weighting. Test 3 is weighted to be highly significant (50) while Test 1 is very lightly weighted (10).

Tests					
	A	B	C	D	Weight
Test 1	0.2				10
Test 2		10			30
Test 3			0.3	5	50

## The algorithms in practice

### Metric Scores

The tool applies metric functions to the metrics to give metric scores. This table shows the scores given to the sample tenders from the first table in this section:

- Metric A is derived by dividing one by the number of bids
- Metric B is derived by dividing the mean price by standard deviation
- Metric C is derived by dividing 100 by the word count
- Metric D is derived by dividing 100 by the Benford’s law score.

Tender Metric Scores				
	Metric A	Metric B	Metric C	Metric D
tender1	0.2	10	0.33	5.00
tender2	1	5	0.02	1.11

### Tests applied

The tool runs algorithmic tests on the metrics. These tests consist of two parts: checking to see if a metric score or metric scores for a tender exceed a set threshold and weighting the tests (see above).

The chart shows the tests being applied to the metric scores for sample Tender 1.

Tests					
	A	B	C	D	Weight
Test 1	0.2				10
Test 2		10			30
Test 3			0.3	5	50

Sometimes tests can involve more than one metric. For example, Test 3 combines Metric C (word count) and Metric D (Benford’s law). This is particularly useful if a combination of certain factors are considered to be a particularly useful indication of potential cartel behaviour.

## Outputs

Each tender is given a pass or fail result for each test and an overall suspicion score. High scores are more suspicious and show where further investigation might be needed.

This chart shows the results of the three tests for the two sample tenders:

- Tender 1 has failed two of the three tests and has a suspicion score of 80. Therefore, it would warrant further investigation.
- In contrast, Tender 2 has passed two of the three tests and has a suspicion score of 10, so it would not require further testing.

Tender Test Scores				
	Test 1	Test 2	Test 3	Suspicion
tender1	pass	fail	fail	80
tender2	fail	pass	pass	10

## Further Amendments

The user can amend the tool to reflect the results of further testing and other investigations into cartel activity.

They can add new metrics, change the weighting for tests, add new tests, adjust the threshold for passing/failing tests, and create tests using more than one metric.

## Using the tool

The tool has been developed as a set of algorithms that can be used on processed tender and bid data to detect indications of cartel activity. A guide to the code for developers as a .readme file is also included.

# Recommendations

## Introduction

These Spend Network recommendations highlight how the tool could be further developed to become a more useful resource for checking tenders for potential cartel activity.

## Recommendation 1: Continue testing and calibrating the tool using more data

To reduce false positives and calibrate the tool in different markets, a greater volume of data is required. The pricing in some areas will be markedly different from pricing in others without this being suspicious, so more understanding of how to set thresholds in different regions is needed. Similarly, some markets will vary in how they score on certain metrics without this being suspicious. For example, aerospace will have a smaller number of suppliers than web design. The tool is variable to take account of these differences, but we will need data on each market to calibrate these metrics.

Data needed to calibrate the tool:

- More councils from a wider range of regions
- More years worth of data
- More markets, expanding our research away from construction and allowing us to
- More construction data to calibrate our current tests
- Cartel data:
  - From existing CMA cartel cases or cases investigated by global competition authorities
  - Test that the tool can identify cartel activity and calibrate it if not.

Accessing data needs to be improved to ensure we can gather this extra data. Using an online repository like Google Drive is the easiest method, although using an Iron Key could also be simple provided there are no problems with postage. Providing access to a tender portal transfers the data gathering work to Spend Network or the CMA, but this was also effective. Sending the data through encrypted emails was time-consuming for both the partners and Spend Network and is not viable for large volumes of data.

## Recommendation 2: Resolve legal issues

Have a proven legal framework that allows councils confidence that when they pass over the data that it will be treated as confidential:

- Option 1 - Establish a robust agreement at an early stage of the data gathering process that the data will be treated as commercial in confidence.
- Option 2 - Never take ownership of the data. Each organisation puts up the data in a form that can be processed by the CMA without taking ownership of it. Councils run the algorithms in their own secure repositories. The anonymised scores would be sent to the CMA and tenders will be flagged if suspicious. At this point, the system suggests the council check the tender. The only data stored away from councils

would be the anonymised scores.

### **Recommendation 3: Create a platform to collect and process the data**

This would enable us to handle the larger volumes of data in a much more efficient way as the manual effort involved in processing data for this project took a significant amount of Spend Network resource.

- A distributed computer model (meaning different software system components shared across multiple computers) would be a useful platform on which to process this data.
- Use an agreed data standard to make it easier to compare and process files. Spend Network recommends the Open Contracting Data Standard, which is backed by the UN and GDS and is gaining traction in wider Government activities at the moment. CMA support for this initiative would be timely and welcome. An agreed data standard would reduce the manual effort required to process files and extract the metrics to analyse from them. <http://standard.open-contracting.org>

### **Recommendation 4: Collaborate with other interested parties**

Spend Network would be happy to host a workshop with any other interested organisations within Government. Entities that have already expressed an interest include the National Fraud Agency and DCLG. We believe expanding the scope of this approach will enable us to improve the methodology and access more data while finding new uses for the tool to benefit Government.

## Concluding Remarks

This project successfully proved algorithmic tests on tender data could be used to identify suspicious signs in tenders. It is very likely that this could be used to detect cartel activity.

Spend Network gathered and processed data from partner organisations and worked with experts from the CMA and these partners to devise, develop, implement and hone a selection of metrics that are indicative of cartel behaviour.

Spend Network developed a tool which tests metrics on tenders and assigns a suspicion score. This tool is configurable so it can be used in different regions and markets. However, the tool needs to be calibrated with a wider range of data, including cartel data, before it can be a useful means of identifying bid-rigging.

The future of this methodology requires a legal solution to the problems Spend Network encountered when accessing data. A platform and data standard must be adopted to handle this data once it is accessed. Despite this, there is great potential for this tool to be used, not only by the CMA, but also by fraud and competition agencies nationally and globally.