STANFORD
COMPUTATIONAL
ANTITRUST

IMPLEMENTATION SURVEY

The Adoption of Computational Antitrust by Agencies: 2021 Report

Thibault Schrepel* & Teodora Groza** (editors)

Abstract. In the first quarter of 2022, the Stanford Computational Antitrust project team invited the partnering antitrust agencies to share their advances in implementing computational tools. Here are the results of the survey.

* **Thibault Schrepel** is Faculty Affiliate at Stanford University CodeX Center (creator of the Stanford Computational Antitrust project), Associate Professor at VU Amsterdam, and Invited Professor at Sciences Po Paris & University of Paris 1 Panthéon-Sorbonne.

** **Teodora Groza** is Ph.D. student at Sciences Po Paris.

Argentina
National Commission for the Defence of Competition

In recent years, digitalization and information technology as a whole have proved to be the cornerstone of modern life. Noteworthy, the current Covid-19 pandemic has boosted the need to update all computational tools for every field.

As the Computational Antitrust Project highlights, "legal informatics and computational law can foster the automation of antitrust procedures and improve antitrust analysis more generally."[1] In that regard, the National Commission for the Defence of Competition (CNDC, for its Spanish acronym) of Argentina has been working on developing an IT system upgrade and aims to move towards fully digitalized procedures.

The first step of this transformation was taken in 2017 when all paper files became electronic records and started to run in a system called *Electronic Document Management* (GDE, for its Spanish acronym). Implementing this system was the beginning of a significant change that was followed by other critical processes.

Indeed, an IT system called *Remote Procedures* (TAD, for its Spanish acronym) was one of those key processes. Electronic records in GDE meant an important development, but every firm's submission was still on paper, and scanning and uploading had to be done by the Commission's personnel. Therefore, the TAD completed the gap by allowing companies to upload digital submissions with their company identification number — CUIT (national tax identification). Filings made through TAD get linked to GDE files. Thus, CNDC's GDE users can immediately see the submission uploaded by the TAD system. On the one hand, this system represents an improvement for the CNDC because processing times have shortened. On the other hand, companies can make submissions more easily and at any convenient time. Moreover, it is easier for the parties to process the files through TAD and check on any case file development as many times as needed.

To complement this, an IT system exclusively designed for the Commission staff was developed. During 2020-2021, CNDC chose to restructure the Registry Directorate where all procedures are submitted. This rearrangement had the main objective of enhancing the registry and the processing of the records dealt with by the Commission. This new IT system that allows the management of files, records, and CNDC's reports is called *Mordelon*.

The *Mordelon* system is comprised by:

1. An archive module that contains the record of all saved files and their exact location, with a search engine that immediately allows users to know that location;
2. A (legal and administrative) document and file management module that is integrated with an administration and information upload interface used by the registry area;

[1] CodeX, The Stanford Centre for Legal Informatics, Computational Antitrust Project, https://perma.cc/9S6H-UG2V.

3. A management interface for "instructing-users," which allows legal and economic analysts to know the state of progress of each of the files assigned to them;
4. A local control interface for "directors-users," which allows directors to know all the files assigned to their area, their status, and the users interacting with them, and;
5. A general control interface for the "authorities-users," which allows commissioners to know all the files assigned to each Directorate of the CNDC, their state of progress, and the users who interact with them.

Control users ("directors" and "authorities") have an overview that provides them with real-time reports on:

- Rulings handed down in the last seven days and communicated to the CNDC by the courts that have dealt with appeals or complaints related to the agency's files;
- Case files that are due to expire in the next fifteen days and therefore require urgent action on the part of the instructor;
- Submissions made during the day by firms in ongoing cases;
- Relevant documents that were signed over the last five days (decisions or resolutions);
- Documents that had any developments over the day and;
- Files initiated at the front desk over the previous seven days.

Control users are also provided with a dashboard that allows them to access specific information:

- The number of cases in process, classified into those with a final decision and those that are still being processed;
- The number of files, classified according to the stage of processing;
- The timeline of each file, with details of the stages it went through during its processing and the respective dates and duration of each stage;
- The record of signed documents, classified into the type of act;
- The register of documents to be signed by the Secretariat of Domestic Trade, which is updated weekly and provides information on their progress;
- The register of documents to be signed by the CNDC, which is updated daily and includes information on their progress;
- The register of documents in circulation and pending signature;
- The register of the number of times that the parties viewed each case file, and;
- The register the official notices issued within the framework of each file and their status in real-time.

To sum up, the upgrade of the *GDE/TAD* systems and the *Mordelon* system has allowed the CNDC to have a report in real-time of fundamental indicators and to work more effectively. Nevertheless, there is still a long way to go towards an integrated digitalized work system, and the CNDC will keep on working in that direction.

Armenia
Competition Protection Commission

The Competition Protection Commission (hereinafter referred to as CPC) was established in 2001 on the basis of the Law on Protection of Economic Competition. The Commission is an autonomous competition authority, independent in the performance of its functions and the exercise of its powers.

The CPC's task is to ensure the freedom of economic activity, free competition, as well as the environment necessary for fair competition and the development of entrepreneurship. It is also entrusted with the function of protecting consumer interests.

As for the computational tools, the CPC is in the process of fundamentally reshaping its activities from traditional to digital. Namely, an electronic platform is being launched that will firstly digitalize the whole internal and external documentation of CPC, and secondly, we will provide access to different state platforms, such as the ones of tax services, cadaster, and state register of legal entities.

The platform algorithms will enable the faster detection and response to anti-competitive behavior of legal entities and state bodies (officials) through the following tools, inter alia:

- The CPC platform has access to a number of state platforms (mentioned above) that provide comprehensive information on the income of legal entities, transactions, etc. Thus, before launching the platform, the CPC gathered information through official correspondence from legal entities and state bodies, and currently, this information is reached with just one click.
- A threshold has been set for transactions. Namely, when transactions exceed the threshold, they will appear on the CPC's platform on a daily basis.
- Of course, while using the platform, human analysis is still needed for understanding whether there is a prohibited concentration or not, however this is a step ahead for facilitating the monitoring activities of CPC.

Australia
Australian Competition & Consumer Commission

Recent & current projects:

- Predictive coding for document review in both investigation and litigation phases of matters (within Nuix Discover);
- Concept clustering to group documents with common themes and identify connections between the cluster groups used to understand the themes across a document population in order to prioritize documents for review as well as to run quality checks on anomalies in the review (within Nuix Discover);
- Use of compare tools to identify the differences in contract versions etc. (within Nuix Discover);
- Combining Natural Language Processing (NLP) techniques with other techniques (vectorisation and clustering) to automatically identify and measure groups of similar documents for investigations;
- Pattern matching to identify phone numbers, credit card numbers across sets of documents;
- Entity recognition using NLP on incoming contacts to the agency, to better use aggregate data on the issue, products and companies of concern;
- Topic modelling on documents and consumer complaints to analyse and visualize the common themes, allowing high-level insight into trends and possible areas for investigation;
- Web-based data collection (web scraping) tools for advanced use to collect evidence, including tools to mimic consumer behavior on websites.

Near-future projects:

- Automated classification of incoming complaints to speed up analysis and response, using natural language processing and a custom ML pipeline;
- Internal predictive ML to estimate time and resourcing for projects and investigations, to support decisions about meeting organisational priorities;
- Online submissions portal with integrated computational support, including data validation.

Brazil
Administrative Council for Economic Defense (CADE)

Background

Any initiative that aims at the implementation of computational tools must deal with underlying issues that are usually neglected.

Such issues are core to the challenges faced by the Brazilian competition authority (Administrative Council for Economic Defense, or CADE) as they involve values and rules computational tools can potentially affect. Therefore, we must look for strategies to ensure the computational models capture our values and rules, understand what we did and, above all, understand what our actual objectives with such tools are. This would be the "alignment problem," one of the most relevant questions of our time: on the one hand, there are technologies of pervasive use and opaque constitution, and, on the other, the values and rules essential to a competition authority's institutional actions. Preventing such misalignment means avoiding potentially deleterious results for the society and the economy.

This is because several of these models have a peculiar characteristic: their continued use, without revision, can reinforce and potentiate preexisting problems – such as biases – as the models are subject to feedback loops. In other words, starting from a reality in which there are inequities – which are usually reflected in the databases adopted by these models – these systems produce results that reflect the original biases, and, as the results will return to the system as input for future analyzes and decisions, crystallizing inequities in a vicious cycle that confirms the same biases. Hence, these models not only mimic but, in the long term, can also change the reality they came from as they consolidate and expand the preexisting inequities.

CADE's experience

The Brazilian competition authority has considerable experience in the development of data mining and screening techniques to enhance the detection of signs of bid-rigging (computational tools for improving detection). The authority's goals with them are initiating new investigations and identifying markets that are more susceptible to collusion. As an example, CADE is currently pursuing an ongoing investigation whose developments drew on the use of screening and data mining techniques entirely.[2]

Even though these tools still strive for accurate results, strategies and heuristics have been developed to avoid the risk of false positives and/or false negatives, such as the use of the bag of screening models combined with data mining.

[2] Based on the screening and data mining results, the authority identified a suspicion of collusion in public tenders, which lead to down raids in 14 companies.

In addition, the authority develops tools aimed at speeding up investigations by automating procedural phases. Based on the analysis of case handlers' workflow vis-à-vis the time spent on each activity, it is possible to see what tools can potentially be automated (computational tools for improving efficiency).

Finally, the use of computational tools for detection activities brought positive externalities to the organization as the tools also collaborated with the work of other areas. For example, CADE's merger units adopted applications developed for research purposes, which in its turn, created new necessities that called for new solutions. That is, the merger units felt the need for additional functionalities to analyze graphs on corporate relationships between companies. Therefore, the tool had to be further improved, bringing about more advances.

For devising these tools, various internal units have engaged in constant dialogue to define common parameters and avoid overlaps between their activities. For instance, (i) the Office of the Superintendent General, through its unit dedicated to the Project Brain, is responsible for conceiving techniques for cartel investigation; (ii) the IT Unit supports the development of initiatives for the automation of internal activities; and (iii) the Department of Economic Studies applies methods to carry out sectoral studies and specialized analyses.

Next steps

CADE is evaluating the development of tools based on artificial intelligence for detecting signs of cartels and improving internal investigation processes.

Besides the mentioned alignment problem, the challenges[3] for detecting signs of cartels are related to the following: AI training data, considering the reduced number of convicted cases; the developed techniques, given the institutional differences of jurisdictions, which can make it difficult to adapt techniques to each local reality; and the transparency of techniques, especially when its results are used for imposing sanctions.

For the second aim, task automation is likely to occur, given the existence of well-known mapped workflows and standardized activities, which would increase significantly the efficiency of the analysis of the work performed, especially the analysis of electronic documents and evidence.

The decision-making process has multiple and complex steps, such as data generation, construction of training bases, the definition of model parameters, tests, and implementation. Hence, it is important to stress that, for the decision-making process, task automation can create biased chains – potentiated at the various links of the AI system — which require extra caution in its implementation. It is important to take into account AI regulatory initiatives, which

[3] It is important to take into account AI regulatory initiatives, which seek to translate, in regulatory terms, the principles and objectives of computational tools.

seek to translate, in regulatory terms, the principles and objectives of computational tools.

Thus, if implemented thoughtlessly, the use of AI-based computational tools can strengthen the alignment problem, as well as raise concerns about the transparency of the techniques.

Colombia
Superintendence of Industry and Commerce

The SIC (Superintendencia de Industria y Comercio) has been developing three projects aimed at improving technological tools and analytical techniques that can help the SIC's investigators and forensic engineers to meet the challenges posed by the large volumes of data produced and consumed by today's markets.

These projects were envisioned as tools that could: enhance the Authority's detection capabilities; increase its investigative efficiency when addressing large volumes of structured and unstructured data; pinpoint suspicious signs, patterns, transactions, and exhibit pricing insights within such data, and better allocate its human resources.

Sabueso

Sabueso was created with the objective of harnessing the data on products available in online commerce (web-based portals), allowing a baseline of information to analyze the historical behavior of products that have been or that are or could be under the scrutiny of the SIC (Superintendencia de Industria y Comercio). It is a price monitoring tool aimed at enhancing the authority's inspection, surveillance, and control functions. Sabueso has four types of information-gathering bots, each of these bots was specifically designed to traverse a supermarket's website, simulating a human's online interaction with the store. This is how information about the products available for sale in each store is collected.

To speed up the collection of information, the work is divided into categories and the system coordinates the launching of a robot hive. Each of the members of this hive oversees the collection of a single category and after finishing its task is replaced by a different bot. Sabueso allows investigators to view descriptive statistics automatically for the products filtered in the selection tab. The descriptive statistics displayed depict the behavior over time of product prices and the variations between samples. In addition to this type of statistical analysis, there are also applications such as the identification of patterns that indicate warning signs in other products and markets.

Sabueso was recently implemented during a VAT-free day –where certain groups of goods are free of VAT to encourage household consumption and economic recovery promoted by the government. The tool monitored and captured the prices of appliances, apparel, and sports supplies sold by a large department store. The tool allows dynamic visualization in dashboards and will provide insights into firms' price strategies during VAT-free days.

Inspector

The highly qualified human talent within the Competition Advocacy Group devotes a significant amount of time to purely operational activities. For example, gathering information on new regulatory projects issued by regulatory entities to analyze later whether such projects may impact free competition. The application developed by the SIC solves the problem by monitoring the regulatory projects published on the websites of state regulators. If changes are detected in the published regulations, the system automatically sends an e-mail to the interested party previously defined to inform about the news.

Sherlock

Sherlock is a project that seeks to support the SIC's investigators in the identification of signs or patterns that suggest potential anticompetitive behaviors with the data available from public procurement processes.

The first stage of development of the project consisted in the creation of a tool that could facilitate the access of investigators to public data available on the web. The purpose is to aid case handlers in the identification of procurement processes with similar characteristics to the one (s) of interest; the bidders within these processes; the contracts awarded by the same entity to the same contractors; relationships or nexus between legal representatives and relevant participants, among others. This first stage of the tool comprised capturing the data, organizing it, and afterward, providing simple descriptive analytics to the investigators. The software cleans, structures, and transforms the collected data and then loads it into a database within the SIC. It also allows for visualization in dashboards with filters, graphs, and comparisons.

Sherlock's second stage, now in the beta phase of development, involves the automation of the search of the abovementioned signs and patterns within the bulk of public procurement data available online. It is based on machine learning technologies and deep machine learning; it relies on artificial intelligence. This second stage will leverage historical data and will analyze and further diagnose information based on indicators already identified by international organizations like the OECD (Organization for Economic Cooperation and Development).

The project will have the capacity of providing informed red flags. It would automatically categorize a particular piece of data into one of the indicators of collusion widely known. This feature will provide predictions useful for the investigators. This would provide a more dynamic Colombia's computational antitrust tools flow of valuable information for the researchers, in addition to greater descriptive insights from the data.

Czechia
Office for the Protection of Competition

The Office for the Protection of Competition (Office) is constantly trying to improve its efficiency in various areas and is therefore also involved in a number of initiatives and projects. These initiatives are described in more detail below.

Project Watson

The Office is currently testing a project, which aims to gather all the information sources that our employees work within a single platform so that it is easier to work with them. This tool should be able to search and find relevant results in many languages and filter unstructured data with the help of AI, so it should be able to evaluate relevance in an intelligent way. Our employees will then have greater comfort in terms of efficiency and speed of searching for resources. This project is not profitable, it only aims to secure more efficient search in resources, but at the same time, some searched pages are crawled by a robot and indexed into our program under certain keywords.

Project Datacros II

The Office has competence in the field of public procurement review and competition law including the punishment of bid-rigging agreements. We are very active and successful in bid-rigging detection and conduct plenty of dawn raids every year. Nevertheless, the Office is constantly trying to improve its methods, find new and more efficient ways of detection by testing different tools, software, etc., and keep up with the increasingly sophisticated practices of companies.

We are trying to create detection methods for use in practice, we have staff who focus directly on this issue. The Office has previously participated in a project in which a program for downloading data from the Public Procurement Journal was created, and seeks to improve and explore other detection options. For example, we are currently focusing on the extension of the input database for bid-rigging detection, the creation of more advanced tests for detection, based on econometric tests and machine learning algorithms.

The above is also the reason why we decided to participate in the project Datacros II. In the case of this project, we see the potential to point out the risks of prohibited agreements in certain areas. We will be able to trace the real owners of the company and also obtain relevant information to reveal especially ownership structures. With this project, the Office can be even more successful in detecting in bid-rigging detection.

The participation of the Office in another project to detect bid-rigging

The Office did participate in a project that led to the creation of a software tool to detect bid-rigging. The system mainly consists of a program created in the R language, which contains several methods that can be used to test data related to public procurement (which should be published in the public procurement journal) and to detect possible indicators indicating the presence of bid-rigging agreements.

The system was used in practice to justify a dawn raid in an investigation of stone sourcing contracts, which was eventually discontinued because although the system found significant disparities between some of the competing gravel and sand supply prices, a further detailed investigation showed that the quality of the required gravel was not sufficiently differentiated in the contracts, which caused different prices, and the system therefore indicated possible collusion.

Subsequently, during further testing and implementation at the Office, it turned out that the detection tool works only with limited data sets, which were not available in the necessary quality for its results to be more widely applicable. Nevertheless, we are currently working with the Ministry of Regional Development on a project aimed at obtaining better data on public procurement contracts (see below). If we manage to obtain more data, then we will continue/renew the bid-rigging detection project.

The participation of the Office in the project Strategy for Electronic Public Procurement for the period 2021-2025, in cooperation with the Ministry of Regional Development

The Office is actively involved in the preparation of the strategy. Its key activities will consist mainly of methodological and awareness-raising support or the provision of know-how in the field of procurement or competition protection. The aim of this project is also to ensure the design of a system aimed at obtaining relevant data to detect bid rigging, as well as compliance with OECD recommendations on public procurement to reduce the risk of bid-rigging.

The project on the building of an EC competition law decisions database

The Office has supported the project of building a fully searchable European Commission competition law decisions database. Its implementation is to be carried out by a team from the University of Leeds led by Professor Konstantinos Stylianou with support from the Oxford Economic and Social Research Council (ESRC).

The participation of the Office in the project CERTIFIX

The offer of cooperation in the use of the platform for the disambiguation of ownership structures has existed since 2017. The electronic tool has long been offered for all the Office's agendas. The developer offered the Office mainly an

unpaid version of the platform, ad hoc disambiguation of ownership structures at the request of the Office, or the development of a separate application only for his needs. The presentation of the project took place at the Office in September 2019 (with the participation of representatives of all departments). Currently, the Department of Methodology and Supervision of Market Power has detailed a material on ownership structure disambiguation and a methodological manual. Cooperation with the (now) TXP Association z.s. is still ongoing.

El Salvador
Competition Superintendency

In some countries, computational antitrust is making its way into the future for the competition agencies, although the use of computational tools[4] as a proactive response for detecting anti-competitive practices a modern is not yet implemented. More experienced competition agencies are just beginning to acquire the technical expertise to develop and use them.[5] Younger agencies rely on technological advances for their reactive response, i.e., dawn raids for investigative procedures.

The Competition Superintendency (Hereinafter SC for its acronym in Spanish) from El Salvador has been active since 2006, and even though it has budget limitations, is always finding new ways to improve and modernize its investigative tools. Its advances in computational antirust are small compared to other jurisdictions, yet still very significant. They relate more to identifying possible anticompetitive practices through traditional ways and using computer forensics than to the use of algorithms to detect illegal practices for now.

Since 2020, the SC has been developing a market intelligence tool that collects information with the intent of guiding and promoting the focused initiation of effective informal investigations, aimed at detecting and punishing cartels through proactive and reactive investigation methods. This proactive tool consists of 4 components: public procurement; markets in general and collecting relevant public information; the use of ICTs for market monitoring; and reactive mechanisms. The tool still uses traditional ways of obtaining and analyzing information, but the aim is to elevate it to a more complex tool, that integrates both traditional and modern tools that can be used in all markets, including the digital ones. Since this tool is still in development, the SC has focused its limited resources to reinforce the reactive tools for combating cartels and other anti-competitive practices.

At the end of 2021 and the beginning of 2022, the CS acquired specific hardware and software to collect, process, and analyze digital evidence, therefore improving data gathering on dawn raids and its following analysis. This equipment, although it may not be considered a proper computational antitrust tool according to theory, is a giant step forward regarding the normal use of technological tools for third-world competition agencies, such as CS.

[4] Defined by Dr. Thibault Schrepel as computer-based problem-solving methods, Thibault Schrepel, *A First Look at Computational Antitrust* (12 September 2021), https://perma.cc/39HZ-CLYN.
[5] Thibault Schrepel, *Computational Antitrust: An Introduction and Research Agenda,* 1 STANFORD COMP. ANTITRUST (2021).

Among the hardware acquired are a TX1 Forensic Imager,[6] which allows reducing the time of making forensic images; an UltraBlock Kit which permits read-only access to data storage devices without compromising the integrity of the data;[7] a Star Tech Drive Eraser,[8] that allows the reuse of the storage devices used by the CS in one case, to make them available for future cases; a Forensic Recovery of Evidence Device (FRED),[9] a workstation that allows connecting the storage media (SATA hard drives, SAS hard drive, USB and CD/DVD/Blu-ray) that contains the digital evidence collected by the SC and includes blockers of writing to guarantee the protection and integrity of the files subject to analysis and thereby preserve the chain of custody of said evidence.

Regarding software, the CS now has an FTK Imager,[10] which is a free tool that allows forensic imaging of computer equipment saves an image of a hard disk in one file or in segments that may be reconstructed later, and helps analyze the digital evidence collected in the raid. The CS also uses Autopsy 4.19.3, an open-source software that facilitates the analysis of digital evidence and allows the identification, selection, and organization of the elements of main interest for the investigation.[11] Some of this equipment was used for the first time in a recent dawn raid made by the CS and has already proven useful for facilitating the search for keywords in the information gathered. It is expected that these tools will improve the CS's ability to obtain and analyze data from companies that are being investigated for anti-competitive practices.

Although the proper use of computational tools like algorithms and artificial intelligence seem far away from the reach of small-countries' competition agencies like the CS, the authority is always looking forward to learning about the recent development in this Antitrust 3.0 era, hoping that the advances in technology and the knowledge that initiatives like the Computational Antirust project provides may help to find ways for these tools to get closer to grasp for the CS, so that it can better fulfill its purpose of protecting, promoting and guaranteeing competition, as required by Salvadorian competition law.

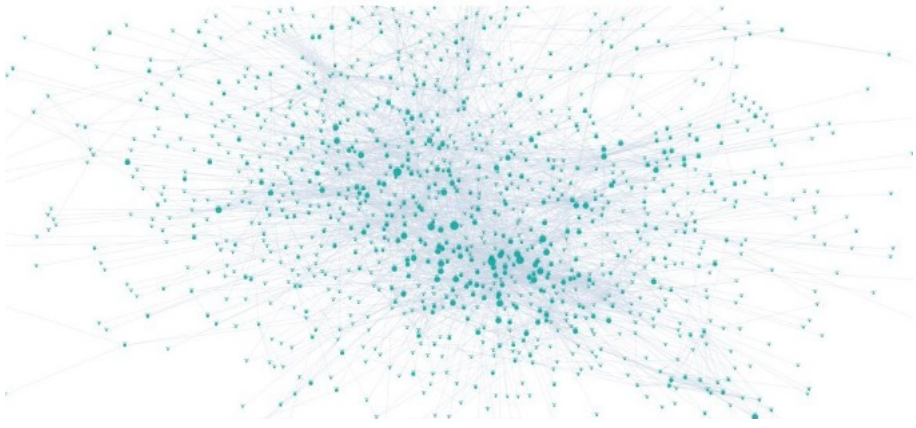
[6] A forensic imaging solution that acquires data faster and from more media types, without sacrificing ease-of-use or portability. opentext Security, https://perma.cc/DU6E-J9JQ.
[7] It is a hardware-based write blocker that allows a USB mass storage device (i.e., flash drives, thumb drives, external USB hard drives, etc.) to be connected to a personal computer for forensic image acquisition and analysis and can guarantee the protection of the data chain of custody. Provides outstanding imaging speed. Digital Intelligence, https://perma.cc/6QC8-KFFA.
[8] This standalone eraser dock erases drives without connecting to a host computer, making it easy to re-use or re-purpose the drive. StarTech.com, https://perma.cc/LHW4-4C37.
[9] FRED systems are high-performance, forensic acquisition, analysis, and processing platforms that set the standard for forensic acquisition and analysis workstations. Digital Intelligence, https://perma.cc/9HMT-CJHT.
[10] Quickly assess electronic evidence by obtaining forensic images of computer data, without making changes to the original evidence. AccessData, https://perma.cc/DAV3-Q8FB.
[11] Autopsy® is a premier end-to-end open-source digital forensics platform. Autopsy, https://perma.cc/T7MP-FGDB.

France
Competition Authority

In 2020, the Autorité decided to create a dedicated digital unit in order to strengthen its resources in digital areas.[12] Launched in September of the same year, the digital economy unit is composed of four people (including two data scientists) and among its objectives, one of them is to develop new digital and computational tools allowing the Autorité to deepen its understanding of the digital sectors and to facilitate the works of the cases handlers. The following presents what the unit has done so far.

The Autorité has collaborated with the ministry of foreign affairs to develop a digital tool called "open terms archives"[13] aimed at following the evolution of the terms of services for the main online service providers. Large digital companies hold today a central position, enabling them, through their ToS, to transform their practices and values into de facto standards which are at the heart of many aspects of our existence and our economies. With this tool, the case handlers will be able to verify the compatibility of these contractual frameworks with national and supranational law and also make sure that they comply with previous commitments. It creates transparency in the practices of digital players.

The Autorité is also engaged in an ongoing project aiming at identifying the quotes between decisions within the FCA in order to unveil the underlying graph (see picture below). For example, the most cited decision is one dating back to 2007 focusing on the market for high-speed Internet access (decision 07 D 33). The Autorité aims to publish an article describing the method used and the early results obtained by the end of the year. The code used to achieve it will also be released on an open-sourced basis at the same time and the final graph will also be made available online.



Picture 1: Overview of the Graph of the Autorité's decisions. The nodes are the Autorité's decisions and the edges the quotes between decisions

[12] The Autorité Creates a Digital Economy Unit, Autorité de la concurrence (January 9, 2020), https://perma.cc/6742-UKUX.
[13] Open Terms Archives, https://perma.cc/C6LE-ZZNE.

The Autorité is also involved in the second phase of the DATACROSS project, which aims to improve the prototype tool assessing corruption risk factors in firms' ownership structure (risks of collusion, corruption, and money laundering in the European single market).[14] It will be useful for cartel detection collaboration with other agencies. In parallel, the Autorité has also begun working on its own detection tool of collusion in public procurements based on the open-access databases available (DECP, BOAMP, INPI, etc.) combined with in-house indicators (to be builded).

Furthermore, the Autorité proposes on its website an interactive map which geolocalizes more than 300 companies fined between 2009 and 2020.[15] Crafted with open and free solutions, it aims at presenting in an innovative and interactive way the work carried out by the Autorité. Further, the underlying data is made available at Data.gouv.fr so that it can be reused in the framework of new projects.[16]

Finally, the digital economy unit is currently developing a variety of automated tools for ongoing cases investigated by the competition units within the Autorité. We will share the technology we have developed for those cases in due course.

Picture 2: Glimpse of the interactive map, centered here on Europe

[14] DataCros II Kick-off Meeting, Transcrime (Feb 23, 2022), https://perma.cc/L5UZ-57FK.
[15] Key Figures, Autorité de la concurrence https://perma.cc/K4EL-Z8EN.
[16] Entreprises sanctionnées financièrement par l'Autorité de la Concurrence entre 2009 et 2020, data.gouv.fr (Sept 7, 2021), https://perma.cc/MJ8U-Z576.

Greece
Hellenic Competition Commission

The Hellenic Competition Commission ('HCC') in the past two years has made a leap forward regarding its technological-digital structure while it is becoming a leader, among the EU Competition Authorities, in the use of computational means in its investigations, already designing and implementing its own data collection platform. Such an 'institutional upgrade' is necessary in order to adequately meet the challenges of the digital economy and increase its effectiveness through the use of new techniques stemming from computational means. Below we present HCC's initiatives in this field, which include the development of computational tools, the establishment of a dedicated unit, and the dissemination of first-class research and advocacy efforts in this field.[17]

Computational tools for the detection of cartels, excessive and exclusionary pricing - developing the tools

The HCC commissioned a report to develop a screening method to detect anti-competitive practices from the analysis of market data and, in particular prices, taking advantage of new legislation enabling the authority to have mandated access to primary data regarding prices by the main supermarkets in the country, the distribution system for petrol stations, and the Athens central market for vegetables and fruits. This enables the authority to follow daily the level of prices for more than 2000 product codes across the country and to be able to use a time series since January 2020 and for some products a few years earlier.

The screen design took into account the following factors; broad applicability, low implementation cost, and accuracy. With respect to its methodology, the screen proposed was based on standard industrial economics principles and builds on top of existing screening methods to detect cartels and was tested using simulation techniques. A two-step method was followed comprising an initial screen to identify suspicious changes in pricing behavior, followed by a verification step to confirm whether the change is indeed consistent with anti-competitive conduct or whether other market events can explain it. The first step of the method consists of a preliminary analysis of unusual and suspicious price changes using a log diff-in-diff model. The second step of the method consists of a more in-depth industry-specific analysis, which involves estimating a pricing regression using supply- and demand-side variables as regressors and testing for structural breaks in the series. This way, it is only necessary to carry out a more complex analysis and to collect non-price data in case of suspected anti-competitive behavior during the previous stage. With respect to data requirements, the first step of the screen requires the collection of pricing data in a panel format. The second step of the method, apart from relying on pricing data, also requires the collection of supply- and demand-

[17] For a detailed account of the HCC's engagement with computational tools as well as its broader ambition to explore ways the implementation of these new technologies may change the direction of competition law see its "inception" report entitled "Computational Competition Law and Economics - An Inception Report" https://perma.cc/P9G6-6Z44.

side data to use as controls. The proposed method has also successfully identified changes in competitive behavior in several simulation tests, performing better than alternative versions of the method.

→ HCC Economic Intelligence Platform

The HCC Intelligence Platform is an effort to integrate and keep updated multiple external data sources in a common database schema and provide visualization tools for data exploration and screening. The platform is being hosted within the premises of the HCC and is accessible for users with appropriate access credentials.[18]

The HCC Intelligence Platform allows the identification of the trends in category and product prices, for every firm or multiple firms and the monitoring of price changes day-by-day or week-by-week (time series options).

A more informative view of the price changes per category or product can be taken by using measures such as the percentiles, median, security threshold, and outliers. Each price is used as a unique data point in the analysis (not an average price of the week) (box plot options). The platform also allows for basket plot analysis. The platform provides a two-step screening tool whereby, as a first step, it conducts a preliminary analysis of unusual and suspicious price changes. As a second step, it tests for structural breaks in the price series.

A special mention of the HCC Intelligence Platform was made in the context of a research project on the use of new digital tools in the implementation of competition supported by the United Nations. The Competition Authorities of Brazil, Russia, and South Africa collaborated with the Hellenic Competition Commission in this program. This gave the Hellenic Competition Commission the opportunity to further develop this innovative tool in the future and to promote Greek know-how internationally.

→ Application

The above screening tool was implemented in a number of investigations during the first months of the Covid-19 pandemic. In particular, it was employed in the following investigations:

- health and hospital equipment procurement;
- markets for basic food commodities;
- healthcare materials.

[18] Link: https://perma.cc/L6LW-LWQF.

Digital structure "upgrade"

Since October 2020, the HCC has established a forensic IT unit, which is headed by an economist and cooperates with a number of data scientists, who are acting as external experts for the authority. In addition, the HCC has appointed a Chief Economist and a Chief Technology Officer.

Next Steps: Establish a Public Ecosystem for Promoting Investments in Applied Complexity Science and Computational Competition Law and Economics

The HCC is in the process of promoting a network of international academic and competition authorities collaborations in order to establish a new public authority-academia platform for the study of complexity science and the development of corresponding computational tools (e.g., agent-based modeling, advanced simulations) in its economy mapping and enforcement activity, with the aim to develop concrete tools that could be of use by other competition authorities members, in the EU and beyond. The broader theoretical framework of this research program was described in the President of the HCC professor Ioannis Lianos' article published in 2022.

Mexico
Federal Economic Competition Commission

The Market Intelligence Unit (MIU) of the Investigative Authority (IA) was launched in 2014 as part of the efforts of the renewed Mexican Federal Economic Competition Commission (COFECE) to enhance and strengthen its enforcement capabilities.

Among its main tasks, the Intelligence Unit detects potential anticompetitive conduct and competition issues in the Mexican market. It also supports the Cartel, Unilateral Conducts, and the Regulated Markets Directorates by providing insights resulting from the collection and analysis of key information from markets, as well as undertakings. Furthermore, it conducts the forensic acquisition of digital information during dawn raids and processes economic and digital data from investigations to aid substantiate cases.

The activities of the MIU are inherently based on the use of technology. This is largely the reason behind COFECE's continuous and substantial investment in the expansion of infrastructure, resources, and staff. In this regard, most of the investment has been focused on broadening IT forensic capabilities, expanding computing capacity, and cloud-based infrastructure. In terms of software, the Intelligence Unit uses mainly open-source solutions that facilitate the creation of specific in-house analytical applications (such as analytical algorithms programmed in R or Python). Another important financing element has been destined to obtain secure storage systems that allow for the safe and independent managing of data.

On its part, the Intelligence team relies on highly multidisciplinary profiles, with a staff that varies from 20 to 25 members (which started at four in 2014). Approximately half of them focus on screenings and economic analysis, whereas the rest develop forensics and other functions. Additionally, because of the complexities the area deals with, it is the unit with the highest level of academic degrees in the IA, as almost half of its members hold at least a Master's degree, and above with varied specialties such as Security, Data, Economy, IT and Intelligence.

In this line of ideas, COFECE has sought to develop sufficient capabilities to explore and further understand whether competition concerns that have been referred to in the literature about technology and competition happen in practice. In this regard, since 2020, the MIU has developed a "DataLab project" which uses specialized infrastructure and an arrangement of several tools that facilitate virtual environments, big data operations as well as the design of algorithms to explore the functioning of specific technologies and assess their potential competition impacts.

The "DataLab project" will allow us to understand how disruptive technologies work. For example, it can be used for understanding blockchain, exploring how the algorithms of search engines work, assessing situations of algorithmic pricing, as

well as experimenting with Artificial Intelligence and Machine Learning Techniques. We expect that the early results of the "DataLab project" could become evident as early as 2022, with some preliminary observations that result from the exploitation of these disruptive technologies.

In addition to this exploratory approach, the "DataLab project" is designed to provide other departments of the IA with access to powerful infrastructure and technological solutions to solve additional operational requirements, such as processing large volumes of information collected as part of ongoing investigations, web scraping, text mining or even designing analytical routines and algorithms for complex data-intensive cases.

Before the implementation of the "DataLab project," the Intelligence Directorate also sought to develop tools and projects based on the application of technology to enhance the detection and investigation of potentially anticompetitive conduct. This has included the design of an in-house public procurement screening tool which has also been shared with fellow competition agencies and has yielded preliminary promising results. Moreover, we have also developed tools that facilitate the analysis of energy markets based on the use of georeferencing and data visualization.

It is important to consider that our objective to enhance our capabilities has also been applicable to digital forensics. Since 2014, we have continuously invested in the expansion of our IT forensics infrastructure, tools, and staff, to be able to collect direct evidence of potential anticompetitive behavior. Some of the projects in this field have even included the design of specialized infrastructure arrangements and techniques to acquire remote information, to even exploring the adoption of Artificial Intelligence in forensic analysis by adopting techniques such as predictive coding.

In short, the relevance of technology for competition enforcement has become evident for COFECE. It is not only about being able to understand how disruptive technologies work but also to fulfill our legal mandate effectively and efficiently by drawing from resources and tools that facilitate our activities and enhance our probability of success. We are convinced that competition enforcers must adopt and include technology as part of their activities.

Romania
Competition Council

Preliminary work

- Aggregated list of indicators for bid-rigging (in-house)
- Draft methodology – jointly with a group of external consultants
    - international expert studies, methodological instruments;
    - data collection, cleaning, matching, and processing;
    - indicator refinement;
    - grouping by use case;
    - recommendations.

Bid-rigging – use cases

Categorized by:

- 3 dimensions:
    a) elementary collusion techniques;
    b) forms of rent-sharing;
    c) resulting market structure;
- Various types of indicators related to each dimension (bid info, pricing info, procurement market shares, etc.)
- Different types of collusion strategies have different relevant indicators

E.g., two companies split a market by winning contracts cyclically while submitting losing bids to mimic competition. In this scenario, one could observe the following:

a) a bid price distribution that is significantly different from the ones of a competitive market;
b) stable market shares;
c) stable number of bids.

Grouping by use case - possible outcome of automatic screening: "highest market X-level risk of collusion scheme Y"

Resources – BD project

- Funding – EU-funded project, 10.2 MEUR
- HR – IT, case handlers, Chief economist unit, project management
- External consultants – academics
- External consultants – IT developers
- Hardware infrastructure
- COTS software solutions
- Internal and external data sources
- TIME – 5 yrs. in the making

Sources of data

Internal sources (e.g.):

- Case management database; Price Monitor on fuels and retail; State Aid Register
- Digital archive (more than 20 years of physical archive)
- Non-structured files from document management system
- Non-structured files collected from dawn raids (special scenario)

External sources (e.g.):

- Databases from partner institutions (National Trade Register, National Agency for Fiscal Administration, National e-procurement system, Ministry of Justice, etc.)
- Non-structured files from different sites (through internet)
- Non-structured data from governmental agencies, media, other institutions

Challenges

- Obtaining and managing data
- Identifying external consultants with specific competition knowledge
- Internal know-how
- Automation vs. manual screening
- Refining the algorithms
- Sustainability

Managing data

- Need to explore datasets and standardize data forms
- Company names vary, can be mapped with ID numbers X&Y SRL; X&Y CO SRL; X&Y; X, and Y
- In the address, the word “street” or “boulevard” varies: “ST “/“ST.” / "STR ","STR." / "STRADA”; ""BD. "/"BD " /"BDUL. "/ "BULEVARDUL "/"BULEVARD "/"BUL."/ “BL-DUL “/“BLVD”
- Code was written to automate the cleaning process
- If legal considerations occur – need to document and justify methodology of cleaning and mapping data

Automated vs. manual/custom

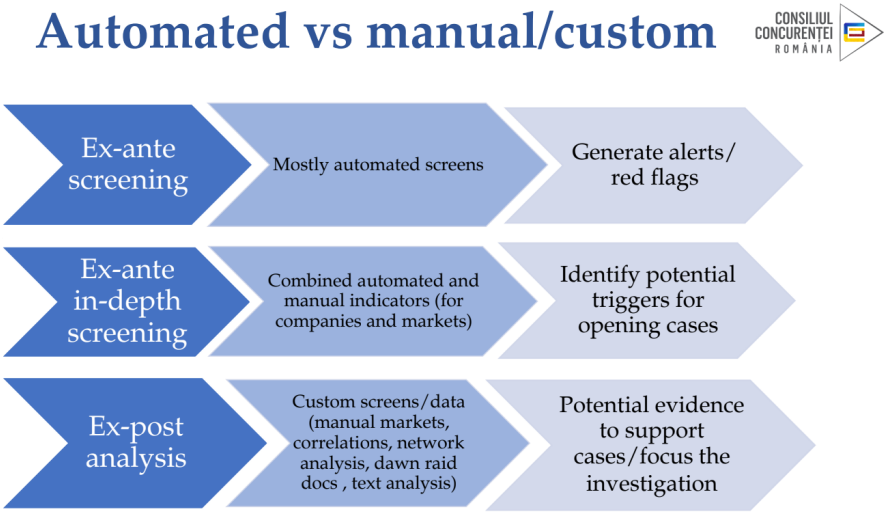
Automatic: indicators based on structured data and automatically calculable (both for screening and specific investigations)

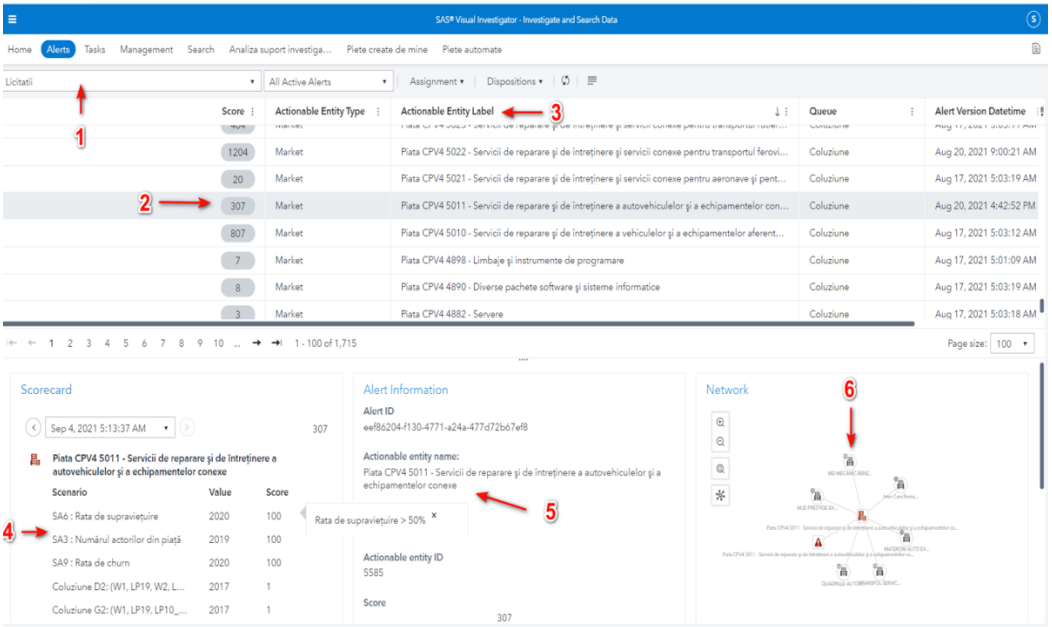
Non-automatic: indicators that require expert assessment

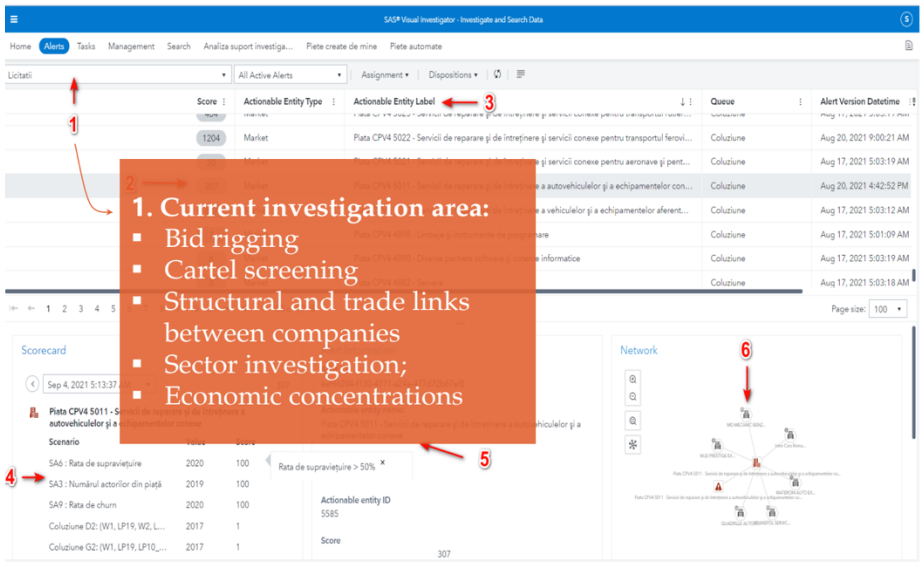
a) unstructured/structured data accessible and regularly processed by the Authority (e.g., market transparency)
b) unstructured/structured data requested from companies on a case-by-case basis (after the start of an investigation/inquiry), or

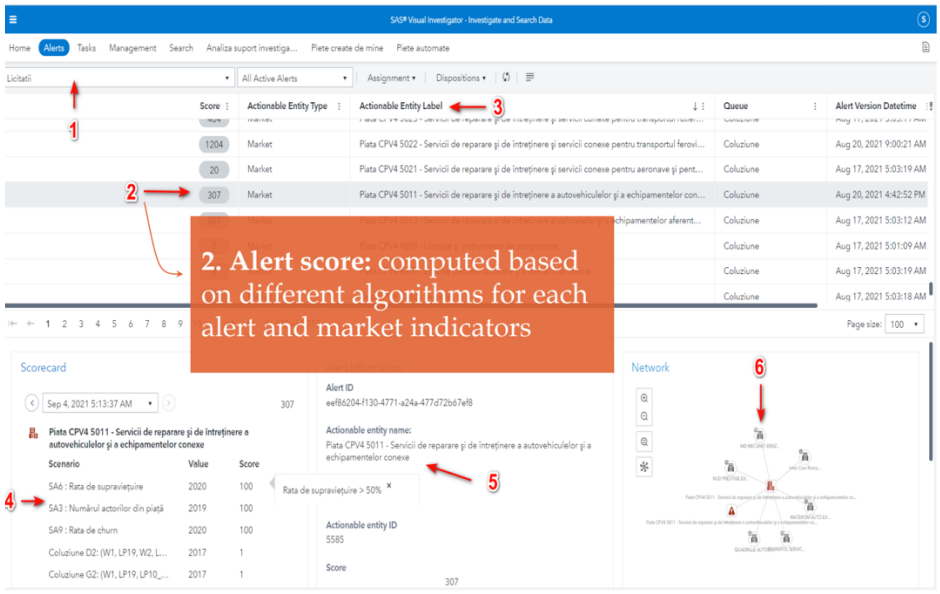
c) unstructured data that is coded manually or algorithmically.

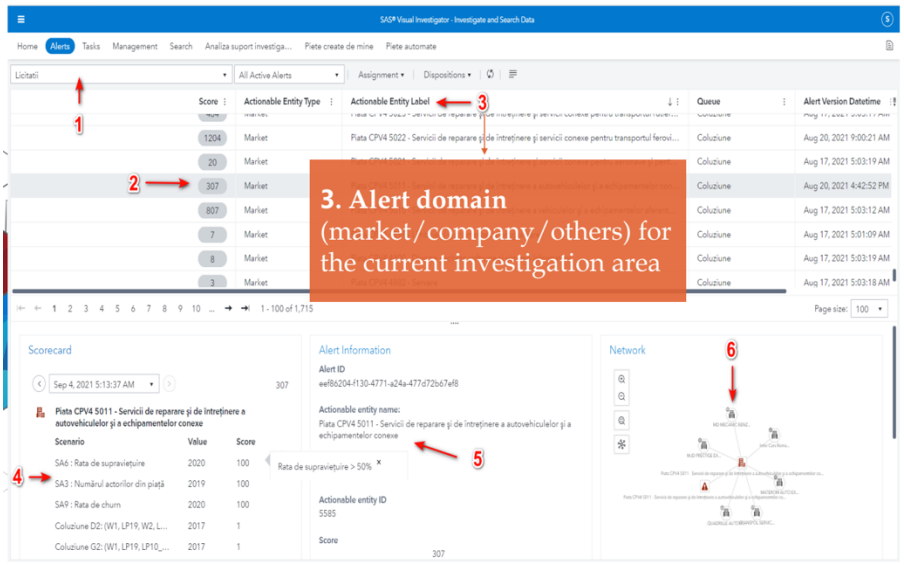
E.g., defining a relevant market - CPV (common procurement vocabulary) vs. customized

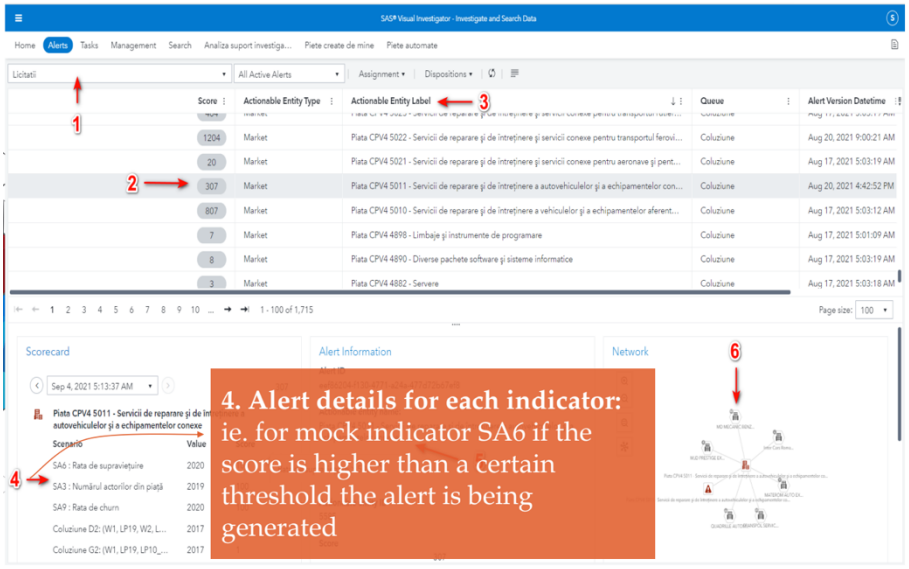


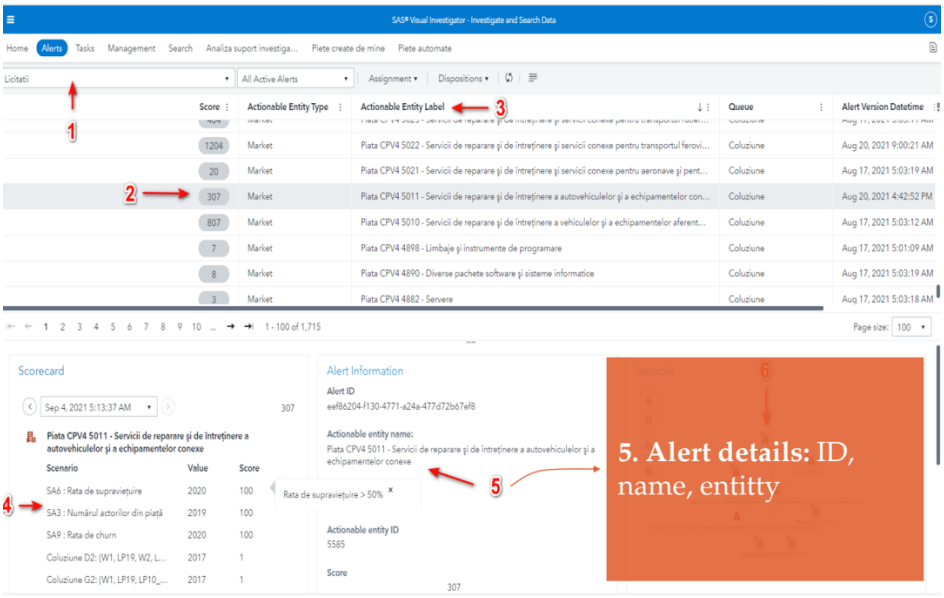


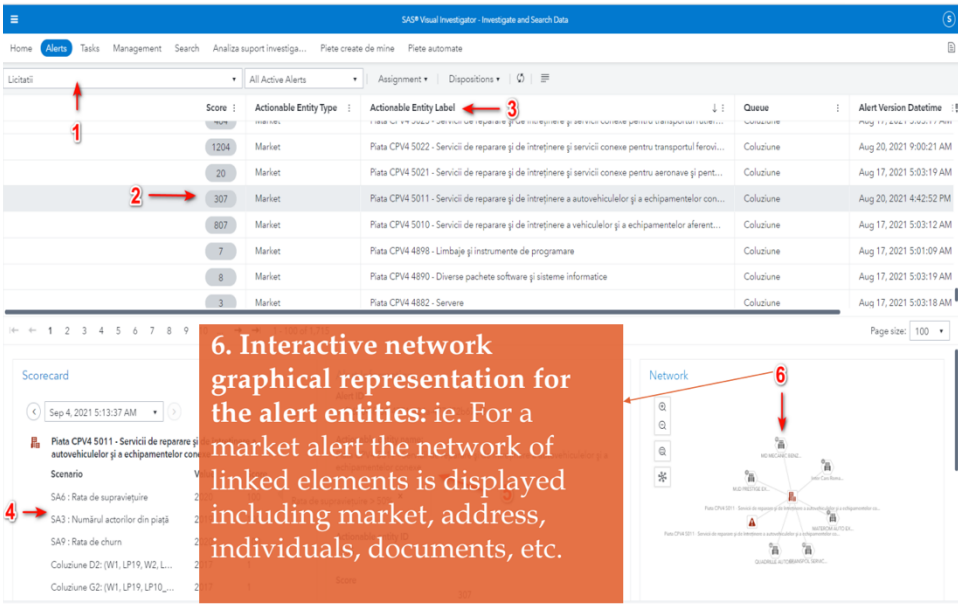

1. Current investigation area:
Bid rigging
Cartel screening
Structural and trade links between companies
Sector investigation;
Economic concentrations

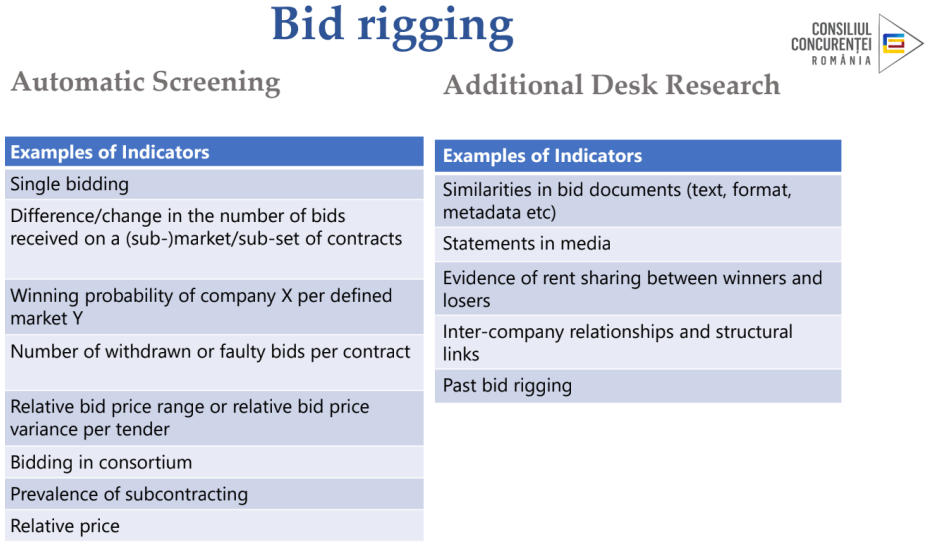

2. Alert score: computed based on different algorithms for each alert and market indicators


3. Alert domain
(market/company/others) for the current investigation area


4. Alert details for each indicator:
ie. for mock indicator SA6 if the score is higher than a certain threshold the alert is being generated


SAS® Visual Investigator - Investigate and Search Data
Home
Alerts
Tasks
Management
Search
Analiza suport investiga...
Piete create de mine
Piete automate
Licitatii
All Active Alerts
Assignment
Dispositions
Score
Actionable Entity Type
Actionable Entity Label
Queue
Alert Version Datetime
1
2
3
1204 Market Piata CPV4 5022 - Servicii de reparare și de întreținere și servicii conexe pentru transportul ferovi... Coluziune Aug 20, 2021 9:00:21 AM
20 Market Piata CPV4 5021 - Servicii de reparare și de întreținere și servicii conexe pentru aeronave și pent... Coluziune Aug 17, 2021 5:03:19 AM
307 Market Piata CPV4 5011 - Servicii de reparare și de întreținere a autovehiculelor și a echipamentelor con... Coluziune Aug 20, 2021 4:42:52 PM
807 Market Piata CPV4 5010 - Servicii de reparare și de întreținere a vehiculelor și a echipamentelor aferent... Coluziune Aug 17, 2021 5:03:12 AM
7 Market Piata CPV4 4898 - Limbaje și instrumente de programare Coluziune Aug 17, 2021 5:01:09 AM
8 Market Piata CPV4 4890 - Diverse pachete software și sisteme informatice Coluziune Aug 17, 2021 5:03:19 AM
3 Market Piata CPV4 4882 - Servere Coluziune Aug 17, 2021 5:03:18 AM
1 - 100 of 1,715
Page size: 100
Scorecard
Sep 4, 2021 5:13:37 AM
307
Piata CPV4 5011 - Servicii de reparare și de întreținere a autovehiculelor și a echipamentelor conexe
Scenario
Value
Score
SA6 : Rata de supraviețuire 2020 100
SA3 : Numărul actorilor din piață 2019 100
SA9 : Rata de churn 2020 100
Coluziune D2: (W1, LP19, W2, L... 2017 1
Coluziune G2: (W1, LP19, LP10_... 2017 1
4
Alert Information
Alert ID
eef86204-f130-4771-a24a-477d72b67ef8
Actionable entity name:
Piata CPV4 5011 - Servicii de reparare și de întreținere a autovehiculelor și a echipamentelor conexe
Rata de supraviețuire > 50%
5
Actionable entity ID
5585
Score
307
Network
6
5. Alert details: ID, name, entitty


SAS® Visual Investigator - Investigate and Search Data
Home
Alerts
Tasks
Management
Search
Analiza suport investiga...
Piete create de mine
Piete automate
Licitatii
All Active Alerts
Assignment
Dispositions
Score
Actionable Entity Type
Actionable Entity Label
Queue
Alert Version Datetime
1
2
3
1204 Market Piata CPV4 5022 - Servicii de reparare și de întreținere și servicii conexe pentru transportul ferovi... Coluziune Aug 20, 2021 9:00:21 AM
20 Market Piata CPV4 5021 - Servicii de reparare și de întreținere și servicii conexe pentru aeronave și pent... Coluziune Aug 17, 2021 5:03:19 AM
307 Market Piata CPV4 5011 - Servicii de reparare și de întreținere a autovehiculelor și a echipamentelor con... Coluziune Aug 20, 2021 4:42:52 PM
807 Market Piata CPV4 5010 - Servicii de reparare și de întreținere a vehiculelor și a echipamentelor aferent... Coluziune Aug 17, 2021 5:03:12 AM
7 Market Piata CPV4 4898 - Limbaje și instrumente de programare Coluziune Aug 17, 2021 5:01:09 AM
8 Market Piata CPV4 4890 - Diverse pachete software și sisteme informatice Coluziune Aug 17, 2021 5:03:19 AM
3 Market Piata CPV4 4882 - Servere Coluziune Aug 17, 2021 5:03:18 AM
Page size: 100
Scorecard
Sep 4, 2021 5:13:37 AM
Scenario
SA6 : Rata de supraviețuire
SA3 : Numărul actorilor din piață
SA9 : Rata de churn
Coluziune D2: (W1, LP19, W2, L...
Coluziune G2: (W1, LP19, LP10_...
4
6. Interactive network graphical representation for the alert entities: ie. For a market alert the network of linked elements is displayed including market, address, individuals, documents, etc.
Network
6



Sustainability

- Managing the IT infrastructure;
- Stimulating internal buy-in from case handlers – pilot projects on past bid-rigging cases to prove utility and showcase the platform;
- Identifying in-house human resources to specialize as data analysts/data scientists/app admins;
- Intergovernmental cloud for increased interoperability.

Lessons learned

- Maintaining and improving data sets is an ongoing exercise - long-term commitment is needed to maintain and update the data sets;
- One of the biggest added-value of the BD platform is the quick access to administrative datasets;
- Analytical needs are too market-specific, which requires further data analysis (combination of quantitative and qualitative desk research) - essential to invest in data analytical skills that go beyond understanding how to use the BD platform;
- Regular implementation of a detection method based on simple screens may hopefully have an additional deterrent effect;
- If trying to "beat" the detection method is still possible, it may increase the coordination costs among cartel participants;
- Ongoing arms race between NCAs and companies;
- Need for a specific cooperation between competition authorities – regular exchange of information on valid and tested screening methodologies.

Singapore
Competition and Consumer Commission

Analytics Tool for Complaints

The Competition and Consumer Commission of Singapore ("CCCS") developed an in-house text analytics tool to analyze large volumes of complaints that are received from the public.

The tool assists CCCS in identifying key industries and thematic issues from the complaints. In particular, it enables CCCS to visualize the underlying time-series trends from complaints and to identify and prioritize a list of industries/issues such as beauty services and renovation. From there, CCCS could take further action by conducting market studies to understand industry developments better or follow up with other government agencies and stakeholders to improve the competitive dynamics of the industry. The tool also reduced the duration needed for CCCS to analyze complaints and allowed case teams to complete analyzing a set of the monthly complaints within a single day, as compared to several weeks with manual filtering and analysis. It also improves the objectivity of the results by reducing the dependence on human judgment which could be highly subjective.

The project involved the use of the R programming language, and software R Studio to create the text analytics tool. Data visualization libraries (such as word clouds) were also used to cluster text around key topics.

Document Similarity and Bid-Rigging Detection Tools

As part of the investigation process for potential bid-rigging cases, investigation teams will generally have to go through the tedious process of reviewing numerous proposal documents submitted by companies for past tenders. Such manual processes are time-consuming and are more prone to human error or lapses.

To address these bottlenecks, the Competition and Consumer Commission of Singapore ("CCCS"), in collaboration with the Government Technology Agency, is building a software tool that automates the process of comparing the similarity of documents at the sentence and document level. The tool employs text analytics techniques such as Longest Common Sequence and Bilingual Evaluation Understudy ("BLEU") to generate similarity scores for sentence and document level comparisons. This allows officers to zoom in on sentences and documents that display high levels of similarity instead of having to comb through large volumes of documents, thereby significantly reducing the time and effort that goes into evidence review, and also minimizing the likelihood of missing out on similar documents through human error.

This document similarity tool can be used in conjunction with CCCS's Bid Rigging Detection Tool ("BRDT") to identify bid rigging behavior. The BRDT is a tool developed in-house that analyses bid prices and bid patterns to flag tenders based on a variety of quantitative indicators that signal suspicious bidding behavior. The document similarity tool is then applied to perform a deep dive into the bid documents submitted to the respective tenders.

Spain
Catalan Competition Authority

The Autoritat Catalana de la Competència (ACCO, Catalan Competition Authority) is an independent and autonomous administrative body of the Government of Catalonia (Spain) whose mission is to safeguard and promote competition in Catalonia.[19]

Why ERICCA

For some time now, the ACCO has noted a significant increase in the number of cases, whether complaints or queries, related to public tenders. Collusion between two or more market operators was one of the main competition issues identified in these tenders. Briefly, economic operators decide to collude when presenting their bids to avoid competition between them to ensure they get a better contract than the one they would have got if they had genuinely competed with each other. By using this strategy, the winning bidder gets a windfall profit which it can then use to compensate the other companies that have fraudulently submitted bids in the public tender. This behavior is known as bid-rigging.

Bid-rigging can take very diverse forms in public procurement, but there are five basic types that may occur individually or in combination:

- **Cover bidding** (also called complementary, courtesy, or symbolic): for example, competitors agree to submit bids that are higher than the bid of the previously chosen winner.
- **Bid suppression:** one or more companies refrain from submitting a bid in the tender process or withdraw their bid so that the previously chosen winner's bid will be accepted (alternatively, firms might choose to submit a joint bid even though they could have bid on their own).
- **Market allocation:** competitors decide to share out the market by geographic areas or contracting authorities.
- **Bid rotation:** companies set up rotation schemes for a number of procedures so that they agree on when they will be the winning bidder.

In response to this, the ACCO has developed a new computer tool called ERICCA (Smart Administrative Procurement Collusion Research Tool).

What ERICCA is and what it does

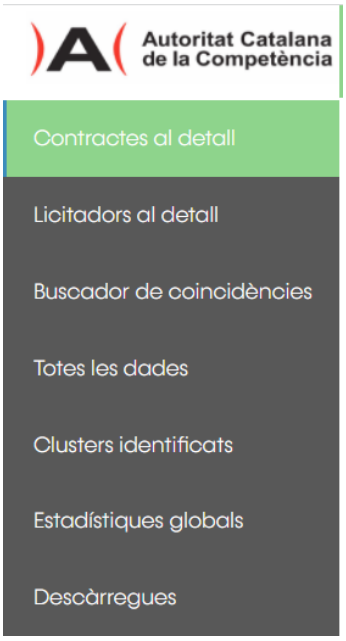
ERICCA is a computer tool that has been developed by the ACCO with two goals:

1. Simplifying the analysis of cases and inquiries in public procurement;
2. Identifying public tenders where participating operators are more likely to have entered into collusive arrangements to submit their bids.

[19] Autoritat Catalana de la Competència, https://perma.cc/9A2Y-U4DY.

How ERICCA does it

The tool features a menu where the user can choose from a number of 7 options:



The first options are designed to ensure better examination of the cases or enquires sent to the ACCO. Selecting the "Contractes al detall" (Detailed contracts) option provides information about a specific contract (participating companies, winning company, tender budget, bid amount, contracting authority, etc.).

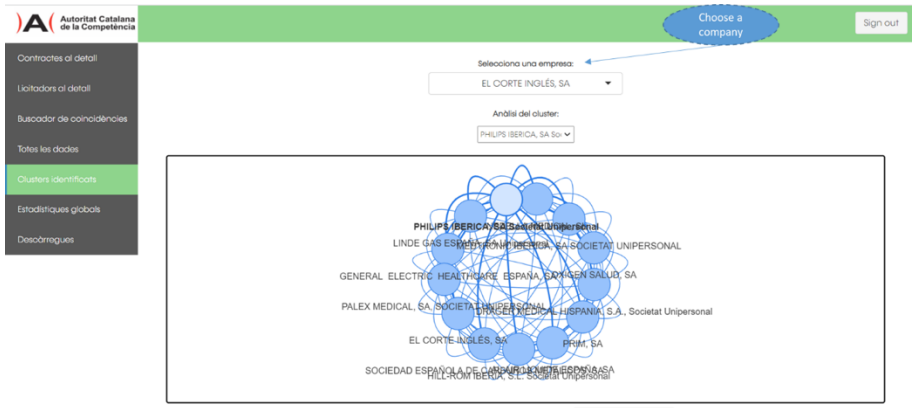
The "Licitadors al detall" (Detailed bidders) option shows all those contracts in which a selected company has participated. It also shows the name of all the companies that it has competed within any of these tenders.

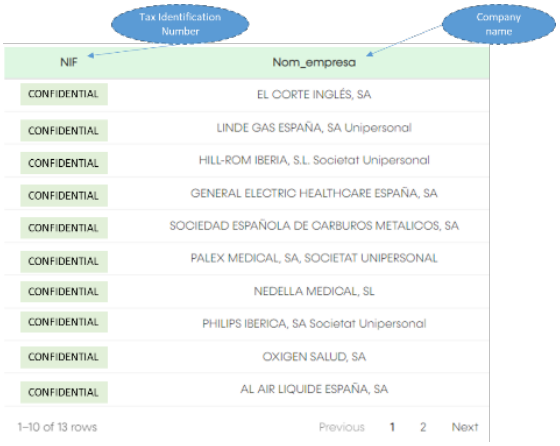
The "Buscador de coincidències" (Match finder) shows all the contracts in which two economic operators who have been selected have taken part.

The "Totes les dades" (All data) option includes all available information.

The "Cluster identificats" (Clusters identified) option is used for proactively identifying potential collusive behaviors. In this case, an Artificial Intelligence algorithm is used. It is an unsupervized machine learning model based on non-Euclidean distances that helps identify clusters between the companies included in the database. This algorithm recognizes companies that are members of the same group, largely because they have taken part in the same tendering procedures, and displays the interrelationships between them as a graphic. Once these clusters have been identified, the tool applies additional filters to detect those where anticompetitive practices are more likely to have taken place. For each cluster, the first filter calculates the average between the tender budget and the award amount. The same filter also calculates the average between the tender budget and the award amount for each company included in the cluster, as well as the number of bids won by each one.

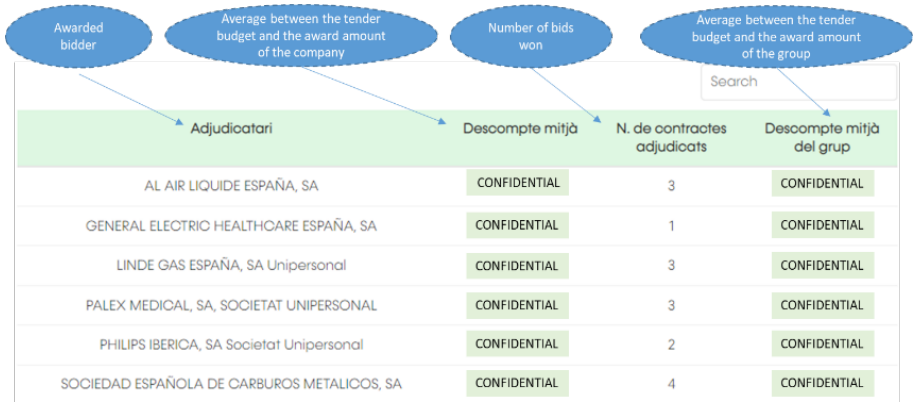
The idea behind applying this filter is to be able to quickly view clusters with differences between the tender budgets and the lowest award amounts, as minimizing the difference between these two variables is one of the main collusion objectives of economic operators. In other words, in the clusters identified that show smaller differences in these variables, it will be more likely in principle that the companies in the cluster have colluded in the tenders in which they participated.





The "Estadístiques Globals" (Global Statistics) option draws on all the information available in the database to show the average discounts (the difference between the tender budget and the bid amount) depending on the selected procedure (open, restricted, negotiation procedure without advertizing, etc.) and the number of participants in the tenders.

Finally, the "Descàrregues" (Downloads) option allows downloading the selected information in CSV format.



Previous problems

In the previous version, ERICCA had to address three major challenges:

I. Web scraping has to be used to obtain baseline data and build up a reliable database for ERICCA. However, due to errors generated by this procedure, the data collected were not accurate, and on some occasions the data readout was wrong.
II. There was no unique identifier for each company since, on the Public Procurement Platform, the name of the same company can be written in several ways. For instance: "Telefónica SL", "Telefónica Móviles SA." and "Telefonica Moviles, S.A.".
III. Information on the financial bids of all the companies was not directly available on the websites but rather only the bid of the winning company.

Next steps

At the present time, with the aim of facilitating the identification of anti-competitive behaviors between the companies, we are creating new filters for the clusters, which include, for example, for each cluster (1) identification of the winning bidders by contracting authority, and (2) identification of the timeframe of the winning bids.

In the future, we will introduce the information corresponding to the cases instructed by the ACCO so that the tool can learn from them and be more precise (this is reinforced learning).

United Kingdom
Competition and Markets Authority

The CMA's Data, Technology and Analytics unit (DaTA unit) – summary of core projects

On 14 June 2022 the CMA published a discussion paper written by Stefan Hunt, *The technology-led transformation of competition and consumer agencies: The Competition and Markets Authority's experience*,[20] which provides more information on the formation of the DaTA unit and the value it has brought to the CMA.

The following note sets out and summarises the core projects which the DaTA unit has undertaken since its inception in 2019.

Digital advertising

The DaTA unit advised on the digital advertising market study which ran for 12 months from July 2019. The DaTA unit helped the case team by:

- Understanding the technology and systems that platforms and adtech providers use, including understanding the role of data in digital advertising and the impacts on competition, privacy and efficiency; and assessing the technical feasibility and effectiveness of potential data-related remedies (access, sharing, controls/restrictions).
- Providing data science and engineering support, including the heavy lifting of designing, coordinating, transferring and analysing 4TB+ worth of datasets from platforms.
- Providing detailed choice architecture analysis to refine interim report proposals.
- Providing support for information requests (e.g., for user testing) to firms, analysis of the responses and development of testing/monitoring proposals for the final report.

Mobile ecosystems market study

In the mobile ecosystems market study, the DaTA unit provided analysis on large data sets from the main Parties as well as insight into mobile ads and Android and Google mobile services. Our behavioural scientists also worked on this study, which covered the core elements of operating systems, app stores, and web browsers on Apple iOS and Google Android. The team analysed how consumers might react to the use of different choice architecture practices on mobile devices and how this can affect competition. This includes evaluating the evidence submitted by stakeholders and academic literature on psychological mechanisms. The team also advised on the design of potential remedies to address the concerns identified.

[20] Stefan Hunt, *The Technology-Led Transformation of Competition and Consumer Agencies: The Competition And Markets Authority's Experience*, CMA (Jun. 14, 2022) https://perma.cc/SNH3-28V8.

Google Privacy Sandbox

The CMA became aware of Google's proposed phasing out of third-party cookies on Chrome, its web browser, and introduction of new technologies that would support targeted advertising while preventing cross-site tracking: Google's Privacy Sandbox. The CMA had a concern that the proposed changes could undermine the ability of publishers to generate revenue through advertising, undermine competition in digital advertising and allow Google to entrench its market power as both a publisher providing advertising inventory and a provider of adtech services. In January 2021, the CMA launched an antitrust case in advance of the changes, working closely with the UK Information Commissioner's Office. Understanding the implications of the changes for the many parties affected has required getting deep into the technologies, the DaTA unit has attended key meetings with the parties and driven the CMA's understanding of the technology. The CMA accepted commitments from Google setting out how the technology will be developed, with the CMA having an ongoing monitoring role.

Meta's gathering and use of advertising data

The DaTA unit advised on a CMA case involving Meta's gathering and use of advertising data for its online classified ads and online dating services. It quickly became clear that understanding the use of data was going to require grappling with how Meta's algorithmic systems function, and so we deployed a data scientist with many years of experience in using machine learning algorithms, including within tech firms. The data scientist made a large difference by enabling the case team to pursue important lines of inquiry that they otherwise would not have.

Online reviews

The DaTA unit has supported the CMA's consumer team with their series of online reviews cases. The first phase of the work focused on the platforms where fake and misleading reviews are traded. This involved meeting, as part of the case team, with eBay and Facebook on their proposals to detect and remove the trading of reviews. The second phase of the work involved the identification of which review businesses to prioritise for in-depth investigation and the subsequent launch of investigations into Google's and Amazon's approaches to identifying and moderating fake and misleading online reviews. As fully embedded case team members, the DaTA unit wrote substantial parts of the detailed requests for information, including requests for considerable data and details of the algorithmic systems that Google and Amazon have built.

The DaTA unit has conducted extensive web scraping on online reviews and used techniques to detect suspicious patterns indicative of fake and misleading reviews. To decide which companies to investigate in-depth, we collected evidence from several different platforms. We used techniques both to detect suspicious patterns ourselves and to ensure that we had a thorough grip on the methods, so in any subsequent investigations we could assess the advantages and drawbacks of firms' approaches.

Case support for mergers

The DaTA unit has provided direct input to mergers cases and performs three different roles as part of a mergers case team:

- First, we develop specific data or technology-focused theories of harm. Two examples are the Amazon/ Deliveroo case and the Meta/Giphy case.
- Second, we aid the case team in understanding the overall market or specific aspects of products in cases where the DaTA unit has preexisting knowledge of the merging sector. Cases here include Google/Looker and Salesforce/Tableau where cloud technology and the variety of different services on the cloud was a major focus.
- Third, we help assess technical remedies, particularly relevant for intellectual property remedies. One example is Tobii/Smartbox where a remedy was proposed that would make the code from Tobii for their products available on an open-source basis.

CMA's Covid Taskforce

The context for our work in this space was that markets were changing rapidly given the shock of lockdown and the introduction of Covid regulations, leading to price spikes and the unexpected cancellation of travel and pre-booked events.

The DaTA Unit built a CMA webform and data pipeline, which was required to be able to receive, process, and analyse over 100,000 complaints from the public about consumer-related issues since the beginning of the Covid-19 pandemic. Data scientists systematically corrected business names through deduplication and matching and also determined the relevant market or sector to which a complaint related using machine learning techniques.

Without the work of the DaTA Unit the CMA would not have had its own real-time complaints information from which it was able to identify issues and would have required a larger number of staff to work on processing information coming through the existing public-facing CMA channels.

Document review

Our Evidence Submission Portal (ESP) is now used for large document submissions to the CMA. The ESP can take in millions of documents, check that they are in the right format, and process them.

ESP exemplifies a situation where there was no existing solution to meet our requirements, and our data engineers and eDiscovery specialists built a product that integrates well with our commercial eDiscovery tool, Nuix Discover. We have reduced the amount of time from submission of documents by the parties to when the merger case team can review the documents from 4 days down to 1 day, which can make a big difference when under time pressure.

In addition, using natural language processing techniques, we have started deploying new capabilities into ESP that provide additional insight on individual documents, helping to speed up the evidence review process. We have deployed three data science enrichments to Nuix. The first is a 'keywords enrichment,' which extracts the most representative words for each document, allowing reviewers to better understand the likely content. The second indicates how likely it is that a document is about competition-related topics (e.g., competitors, prices, market structure, etc.). The third indicates whether a document talks about activity that is within the UK.

Companies Network Project

The DaTA unit is creating a data pipeline which takes records from all registered limited companies from Companies House, the UK's registrar. These data are frequently needed by the CMA for many reasons, including i) getting intelligence on companies with respect to suspected cartel activity, ii) understanding ownership structures in markets, and iii) understanding the state of markets, especially concentration and profitability levels. Ordinarily CMA staff navigate the publicly available search tool and download the data by hand. But this is time-consuming, and the manual nature of the process could lead to errors. The pipeline will regularly take in all the data, clean it, deduplicate it, and make it available in an easy-to-use tool, designed for the CMA's needs (with the pipeline also available more widely). This is a significant engineering challenge, especially as the dataset is reasonably large

The payday lending market investigation

This market investigation concluded by implementing a remedy that lenders needed to put a link to a price comparison website on their webpage. The DaTA unit created code to scrape company websites to check that there was a link, writing to any non-compliant firms and telling them to comply.

DaTA unit's LEDA platform

The DaTA unit uses its own platform LEDA for creating a data lake (a system for storing and accessing raw data) and providing analysts with access to the data lake and to cloud computing services. Our data scientists and engineers need flexible access to coding packages, version control of all code and data (using GitHub) so that we can reproduce all analysis, and the ability to build and control information security and data protection. Working closely with our IT department colleagues, we developed a data infrastructure capable of ingesting, curating, and processing sensitive data at the scale required of a leading competition authority and with modern advanced functionality. In the last three years, we have acquired over 160TB of data across over 130 million objects stored in our cloud-based data lake, at minimal cost and risk. This is important for our quantitative teams, who have excellent alternative employment opportunities and really care about quality of the 'tech stack': they have the tools they need for the job, like the platform, and feel empowered.

Autorenewals cases

The DaTA unit's Behavioural Hub has been working with the CMA's consumer team on a series of cases focusing on autorenewal contracts for McAfee and Norton antivirus products and gaming subscriptions for Microsoft Xbox, Sony, and Nintendo. Consumer behaviour was central to the investigation. The DaTA unit analysed firms' interactions with consumers – their emails, websites, notifications – and the precise elements of the design. Much is known about the effects of this design – defaults, ease of exit, comprehension of contract terms, provision of risk information – and the behavioural insight advisers built up very detailed evidence on the impact of these elements. The team advised on data collection, including what data and information could be requested from firms and how it could be used to support legal arguments. This included advising on how changes introduced by firms could be used to assess the impact on consumers using quasi-experimental methods. As the cases progressed to remedies, the team members were active in assessing the likely impact of the remedies and so what the CMA should accept in terms of undertakings.

Algorithms programme

In January 2021, the DaTA unit published new research on algorithms, showing how they can reduce competition in digital markets and harm consumers if they are misused. The paper aimed to lay out the territory that the CMA would need to cover over the next 5 to 10 years. There was subsequently a panel event which discussed the key issues in more detail.

The DaTA unit has also conducted additional research on algorithmic harms and algorithmic auditing working with other UK regulatory agencies and anticipates publishing a paper later in 2022 on the CMA's practical lessons from cases involving algorithms.

Online choice architecture (OCA) programme

The DaTA unit's Behavioural Hub published a discussion paper and an extensive evidence review in April 2022. The aim of the research was seeking to lay out a roadmap of an increasingly important area. OCA featured prominently in the Furman report and the Stigler Centre report on digital markets. And OCA is currently an important part of many digital cases, e.g., browser defaults on Apple devices, app stores in mobile ecosystems, or privacy settings within Google's Privacy Sandbox.

The first paper outlines the harms that can arise (while noting that OCA can be and often is hugely beneficial) and includes a taxonomy of 21 concerning practices that agencies need to be aware of and alert to. The second paper is a long and detailed review of the practices and cross-cutting themes, which is intended to be a reference document for agencies, providing a trove of evidence to support future cases.

Stanford Computational Antitrust Project Director: Dr. Thibault Schrepel // Editor-in-Chief: Teodora Groza
Editors: Aleksandra Wierzbicka, Glen Williams, Alessia Lercari Bueno, and Kirill Ryabtsev
Academic Outreach Chair: Jan Jakub Przerwa // Executive Operations Chair: Alex Sotropa
Development Chair: Maura Carey

Hosted by Stanford University CodeX Center