STANFORD
COMPUTATIONAL
ANTITRUST

IMPLEMENTATION SURVEY

The Adoption of Computational Antitrust by Agencies: 2nd Annual Report

Thibault Schrepel* & Teodora Groza** (editors)

Abstract. In the first quarter of 2023, the Stanford Computational Antitrust project team invited the partnering antitrust agencies to share their advances in implementing computational tools. Here are the 26 contributions we received.

* Associate Professor at the Vrije Universiteit Amsterdam; Faculty Affiliate at Stanford University CodeX Center (creator of the Stanford Computational Antitrust project); Invited Professor at Sciences Po Paris & University of Paris 1 Panthéon-Sorbonne.

** Ph.D. student at Sciences Po Paris.

Table of Contents

Argentina 57
Armenia 60
Australia 63
Brazil 66
Bulgaria 69
Canada 71
Catalonia 73
Chile 76
Colombia 81
Cyprus 84
Czechia 87
Finland 90
France 92
Greece 94
Kenya 108
Luxembourg 115
Mexico 117
Poland 121
Romania 126
Singapore 130
Slovakia 132
Slovenia 137
Spain 143
Taiwan 148
Türkiye 152
United Kingdom 155

Argentina
National Commission for the Defence of Competition

During 2022, the National Commission for the Defence of Competition ('CNDC') continued to focus on digital transformation initiatives to enhance antitrust analysis, working primarily on updating and improving its information technology systems with the collaboration of all departments involved in their use.

This article reviews different computational tools that are used by the CNDC, starting with *Mordelon*, an internal file processing system through which the competition authority handles all presentations submitted by third parties. It will also examine the recent developments of the Electronic Document Management Platform, a widely used platform designed to manage and keep track of all government-related processes, as well as the efforts to implement the Remote Procedures System to other types of processes besides merger control. Finally, we will also look at the addition of new functionalities and case law search engines to the institutional website.

Mordelon was developed in 2020 during the peak of the Covid-19 pandemic. The system is maintained and run by the Unit of the Registry, which also keeps it up to date and constantly incorporates new features, adapting it to new processing information requirements. The entirety of CNDC's staff currently uses *Mordelon* for recording the third parties file processing. The system also enables case handlers to interact with the Unit of Registry and enter notification instructions to the area´s personnel.

Mordelon is compatible with the *Electronic Document Management Platform* ('GDE'), a platform through which most third-party submissions are formally filed. It has proven to be an excellent ally for the CNDC staff to manage electronic records transparently and to access real-time reports on the procedures that involve the handling of each case. GDE also minimized the requirement for the CNDC to handle loads of documents that must be preserved and protected, enabling the authority to save resources that were allocated for the preservation of paper archives. Moreover, digitalization rendered all the activities that make up each investigation suitable for remote work, allowing each analyst to have a more agile, flexible, and autonomous work regime, with less need for constant intervention of the Unit of Registry.

Regarding the bridge between third parties and the CNDC, it is a role of crucial importance that is being fulfilled by the *Remote Procedures System* ('TaD'). TaD is,

strictly speaking, a feature of GDE, a system that enables companies and individuals to submit digital presentations to the CNDC using their tax identification number. These presentations are linked to the electronic file in the GDE environment after being reviewed by the Unit of Registry personnel. TaD is quite popular among practitioners because, if enabled, it allows them to remotely monitor and inspect the investigation record's contents as they are uploaded to the electronic file without having to visit the competition authority's physical desk.

In the last few years, TaD has been implemented for merger control-related procedures. Given its success, attorneys and law firms frequently interacting with the CNDC constantly request implementing the TaD feature in the investigations of anti-competitive conduct. Yet this poses challenges.

The TaD system permits not only remote documentation submission but also allows the examination of all the documents on the electronic investigation file by an authorized third party. Anti-competitive investigations thus require an adaptation to prevent confidential information from being revealed to all parties that may be involved in the conduct and have access to the record. This scenario of unintentional and improper access to sensitive information is not a significant problem in merger control cases, which is almost always a collaborative procedure between two associated parties and the competition authority.

To solve this obstacle, CNDC and the National Directorate of State Digitalization of the Secretariat of Public Innovation—a unit of the Office of the Chief of Cabinet of Ministers—teamed up to develop a particular procedure through the TaD platform, which would link confidential documents to a reserved electronic file, accessible only to the CNDC case handlers, leaving non-confidential information in the main electronic file, which could be accessed unrestrictedly by the parties under investigation. CNDC personnel are currently testing this approach in a sandbox environment. If successful, CNDC expects to launch TaD for anti-competitive investigations in the coming months once the training process for the staff of the Unit of Registry and the Unit of Anti-competitive Conduct has been completed.

In addition to issues related to procedures and IT systems, it should be noted that the CNDC is also updating and optimizing its website to create a virtual platform that not only serves as a repository to consult case law but also becomes a tool for effective communication between the competition authority and stakeholders.

The Advocacy Unit is currently redesigning the institutional website in order to enhance its accessibility and provide valuable and straightforward content that is also visually appealing. The incorporation of infographics to accompany the press releases of the CNDC's various actions, the constant updating of the website with the agency's latest news, and the systematic incorporation of new opinions, resolutions, working papers, publications, and other documents of interest stand out among these efforts.

The website has also incorporated 'search engines' that make it easier to obtain specific information; there are databases for administrative decisions and case law, as well as a specific search engine for sanctions for 'gun jumping.' These databases allow practitioners and people interested in filtering jurisprudence by various criteria.

To provide an objective measure of effectiveness, the Advocacy Unit prepares an interaction metrics report to evaluate the performance of the institutional website. This monthly report also includes a follow-up of the interaction indicators of LinkedIn, the only social network that the CNDC currently uses. The tool used to analyze the website performance is Google Data Studio. The main parameters observed are the visits to the website and its specific content, the monthly visits, and the accumulated annual variation. As for LinkedIn, the report tracks the number of followers, views, reactions, and user comments on the news items included and the interaction rates of each content. The metrics report allows the CNDC to have an in-depth view of the reach of all material produced and shared and to adjust the communication and advocacy strategy to be increasingly influential and far-reaching.

To conclude, much progress has been made toward digitalized competition enforcement regarding the work inside the agency and the interaction between the agency and the interested public. However, computational antitrust is a challenging subject, and the CNDC will continue upgrading all IT resources to deal with it.

Armenia
Competition Protection Commission

The Competition Protection Commission (hereinafter referred to as 'the Commission') of the Republic of Armenia is an autonomous competition authority that was established in 2001 under the Law on Protection of Economic Competition. The primary objective of the Commission is to ensure the freedom of economic activity, free economic competition, fair competition, and the necessary environment for entrepreneurship development, also entrusted with the function of protecting consumer interests.

In 2021 the Commission started to fundamentally reshape the activities from traditional to electronic by developing an electronic platform that would streamline its internal and external documentation and offer access to different state platforms.

As of March 2022, the electronic platform has been successfully launched. The electronic platform that is called "e-Compete" consists of an official interactive website and an internal case management system, the mutual integration of which enables to collect information from external sources and transfer them to the internal system, as well as to automatically publish information deriving from internal procedures on the Commission's website.

With various full-fledged instruments, the platform allows analyzing and comparing relevant markets, controlling transactions made by legal entities in order to identify undeclared concentrations, controlling the implementation of the Commission's decisions, monitoring prices for a number of goods on a daily basis, etc.

The platform's core operational systems are outlined below, showcasing its robust capabilities and advanced features.

I. Interoperability with electronic platforms of other state institutions

- The electronic platform offers access to a number of state institutions platforms, such as the state revenue committee, cadaster committee, state register of legal entities, etc. This advanced feature has enabled the

Commission to seamlessly integrate its operations with other state entities, resulting in a more streamlined and efficient regulatory process.

- Through web services the Commission receives different kind of information, such as: shareholders of the legal entity, participation in other companies, group of persons, areas of activity, amounts of assets and revenues or personal information about physical persons, etc.
- The algorithms auto-analyze datasets on legal entities from different state authorities, and flag transactions that are most likely to contain possible violations of the law in the context of merger control.

II. E-filing and registration in the Commission's electronic platform

- Stakeholders can now easily register and submit their documents to the Commission electronically, in a matter of clicks, through the e-filing system eliminating the need for physical submission processes. This has effectively minimized the time and costs associated with traditional paper-based processes, making it a more efficient and cost-effective option for all stakeholders involved.
- The electronic platform fully digitized the Commission's internal document flow and processing, spanning from the initiation of the administrative proceedings to the crucial decision-making process, as well as appeals.
- All templates of documents for submitting to the Commission are available in the system in electronic forms, and submission of documents is simplified and does not require electronic signature. Moreover, the process of document submission has been streamlined and simplified to such an extent that electronic signatures are no longer necessary, making it easier than ever for stakeholders to submit their documents in a timely manner.
- All decisions made by the Commission have been digitized, and the electronic platform generates statistical reports based on the data of these decisions.

III. Monitoring of prices for everyday transactions with platform algorithms

- The electronic platform provides an efficient and effective monitoring of the prices of selected goods. Namely, the Commission has access to the prices of goods and services collected daily from the databases of State Revenue

Committee, which allows it to closely track and analyze the prices of various goods in real-time.

- The platform's algorithms enable the Commission to monitor changes in prices and detect any instances of price fixing or other anticompetitive behavior.
- The Commission automatically publishes information on its website, providing transparency and accessibility to the public.
- Consumers can easily access this information by selecting the specific type of goods they are interested in or searching for a specific seller.
- To conclude, although human analysis remains an indispensable component of our daily processes, the implementation of the electronic platform has vastly improved the Commission's monitoring capabilities, enabling us to swiftly and efficiently identify potential instances of anticompetitive behavior. The Commission remains committed to exploring and leveraging cutting-edge computational tools in its daily processes firmly believing that by embracing the latest technological advancements, we can better safeguard fair competition and promote the interests of consumers, paving the way for a more equitable and prosperous future.

Australia
Australian Competition & Consumer Commission

The Australian Competition and Consumer Commission ('ACCC') is an independent statutory authority established in 1995. Its role is to administer and enforce the *Competition and Consumer Act 2010* and other legislation, promoting competition, fair trading, and regulating national infrastructure for the benefit of all Australians.

The ACCC focusses on taking action that promotes the proper functioning of Australian markets, protects competition, improves consumer welfare and stops conduct that is anti-competitive or harmful to consumers.

The proliferation of data across the economy brings challenges but also opportunities. Key among those is the potential to detect competition and consumer issues sooner than before. As the ACCC expands its monitoring capabilities, and it refines methods to detect structural and behavioral anomalies, it increases the likelihood of detecting new issues as they emerge. The ACCC considers proactive detection to be a fundamental component of its growing strategic capability.

The Strategic Data Analysis Unit (SDAU) was established in 2017 to provide analytical advice and support to help the ACCC's market inquiry work. In 2021, SDAU was combined with several other teams to form the Data & Intelligence Branch. Broadly speaking, the Data & Intelligence Branch helps other teams in the agency obtain and make sense of vast amounts of information for the purpose of making informed decisions. In addition to SDAU, the Data & Intelligence Branch includes: the Infocentre, the Intelligence Team and the Legal Technology Services Unit. The Data & Intelligence Branch works closely with the Information Management & Technology Services Branch to design technical solutions that help the agency address emerging challenges.

Below is a list of recent and current, and possible future, projects that the ACCC is working on.

I. Recent & current projects

A. Using Nuix Discover

- Ongoing use of predictive coding for document review in both investigation and litigation phases of matters;
- 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;
- Trialing the creation of common terms across merger investigations and to use these as search term families as a starting point in prioritizing early case assessment and document review;
- Use of automated audio and video transcription tools to reduce time and cost and increase functionality in reviewing audio/video files;
- Use of compare tools to identify the differences in contract versions.

B. Other Artificial Intelligence and Machine Learning (and related) projects

- Using Natural Language Processing (NLP) techniques to automatically identify and measure groups of similar documents for investigations;
- Pattern-matching to identify entities (e.g. 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 issues, products and companies of concern
- Topic modeling on documents and consumer complaints to analyze and visualize the common themes, allowing high-level insight into trends and possible areas for investigation;
- Web-scraping to collect information, including tools that mimic consumer behavior on websites;
- Online submissions portal with integrated computational support, including data validation;
- Cartel screening to detect possible bid-rigging in procurement;
- Automated searches of business records to identity persons of interests in cartel investigations.

II. Possible future projects

- Automated classification of incoming reports 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 organizational priorities
- 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.

III. 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)

I. Context

Since its development, *Project Cérebro* emerged as an initiative to build partnerships between CADE and other investigative bodies.[1] However, with regular changing staff—within the project scope and in partnering agencies and government bodies—the challenges to recreate bonds and (re)build trust amongst partners became apparent.

The literature points to possible harmful effects in contexts in which teams responsible for developing new technologies operate uncoordinatedly without established parameters or information on the evolution of competitors.[2] This race for creating computational tools can produce ineffective results, particularly concerning undesirable effects on the accuracy and alignment between objectives and values.

These findings can be extended to the government sector, mainly when the performance of agencies overlaps, that is, in cases where several teams in the government sector develop computational tools to solve the same problem.[3] In this regard, consolidating effective and lasting partnerships can be a strategy consisting of reduced costs and significant impacts.

II. Recent strategy

The year 2023 marks the revival of inter-institutional cooperation in the Brazilian federal government. Regarding *Project Cérebro*, the opportunity means the

[1] This text was written by Felipe Roquete, Coordinator-General of Antitrust Analysis of the Office of the Superintendent-General of CADE.

[2] Stuart Armstrong, Nick Bostrom and Carl Shulman, *Racing to the Precipice: a Model of Artificial Intelligence Development* 31 AI & SOC. 201-206 (2016).

[3] Regarding the cartel investigation cases in Brazil. In addition to the antitrust agency, the Prosecution Services and the police can also investigate (*i.e.* criminal investigation) cartels within the private sector. As to cartels in public procurements, besides the already mentioned organizations, competence also lies in the different Offices of the Comptroller-General and Courts of Accounts.

possibility of strengthening partnerships, aiming at an efficient allocation of human and organizational resources.[4]

Therefore, to encourage inter-institutional coordination, CADE has continuously sought to develop long-term collaborations with organizations that have experience using computational tools for data analysis and investigations.

Thus, we avoid federal agencies entering a race to develop innovative solutions, which would implicate possible overlaps, inefficiencies, and redundancies.

The priorities in the public sector are to (i) avoid redundancy of the efforts of agencies; (ii) share techniques and models that enable other agencies, with competence to investigate collusive conduct, to carry out their activities and amplify the dissuasive effect of the tools elaborated by CADE; (iii) internalize tools, developed by other federal agencies, that provide useful inputs to the core intelligence and investigative activities.[5]

Hence, the partnerships with the Office of the Comptroller-General of Brazil and the Brazilian Court of Accounts—which possess competencies to investigate conduct related to federal procurements—were strengthened in 2023. Not only do these bodies develop computational tools for analyzing procurements, but they also have teams dedicated to data science and innovation.

That means the strengthening of the partnership between CADE and those agencies aims at ensuring that the 'race' for innovation and development of computational tools happens effectively and rationally, without redundant efforts of the government bodies, respecting each organization's competency and guaranteeing to bring into use the expertise of teams where it has more relevance.

[4] Although antitrust enforcement is considered cost-effective. *See* John M Connor & R H Lande, *The Prevalence and Injuriousness of Cartels Worldwide*, *in* ELGAR RESEARCH HANDBOOK ON CARTELS (Peter Whelan ed., 2023).

[5] J D Jaspers, *Strong by Concealment? How Secrecy, Trust, and Social Embeddedness Facilitate Corporate Crime* 73 CRIME, L. & SOC. CHANGE 55–72 (2020). In describing the inherent characteristics of collusive agreements and the similar difficulties imposed on their investigation, the author highlights the relevance of the dissuasive effect of the antitrust authority's performance.

III. Results

Since January 2023, the Project has developed new means to detect cartels in procurements, making possible the automation and standardization of the analyses carried out in investigations.

Once implemented in audit procedures of the Office of the Comptroller-General of Brazil and the Brazilian Court of Accounts, we expect to have a broader level of detection of anomalies and evidence of cartels in public procurements.

The partnerships will not only allow for improving violation detection but also have better dissuasive effects regarding agency performance.

Bulgaria
Bulgarian Commission on Protection of Competition

The Bulgarian Commission on Protection of Competition ('CPC') currently does not use any sophisticated computational tools for processing internal or external data. While the Commission is constantly trying to improve its efficiency in various areas, there are still several challenges that the introduction of machine learning systems creates. Therefore the Commission is involved in initiatives and projects on the topic and follows the current trends in order to adopt the most optimal system for its needs.

The Commission has solid expertise with bid rigging cases. Considering the best practices of other NCAs and international organizations, such as the OECD, the CPC has developed its own Guidelines for fighting bid rigging. The new Guidelines have a list of the most common red flags, indicated by many authorities, most of which red flags are applicable to machine learning and automated search engines.

The CPC has also adopted internal Methodology for bid rigging screening based on economic and statistical analysis and currently an ongoing project for an internal document is taking place. The screening methodology is applied during the preliminary investigation. The results of the screening can be used for establishing fluctuations in the competitive process during the tender procedures. The analysis is aimed at comparison of the winning and losing bids in search for relations between them.

The Bulgarian Commission enjoys a wide variety of online and easily available databases, updated both by public and private entities. These sources of information allow the CPC to make fast, up-to-date and reliable enquiries, without the need to send formal requests to different bodies. While the Commission relies mainly for the information extraction from the open-source databases, in many cases the options to filter and/or sort the queried data are rich and powerful enough. This makes the manual processing of information sufficient for the ongoing investigation without the need of a designated computational toolkit.

These databases include registries such as:

- Trade registry with full access to the company profiles and documents, current Articles of association, representation rights, links between the shareholders of separate companies, etc.

- Digital Public Tenders Platform where all public procurement notices are announced, and all corresponding documents are uploaded.
- Food Prices Information Portal – a free online comparison tool for basic food prices, such as fruits, vegetables, bread, milk products, meat, rice, etc.
- Fuel prices tool – a free online comparison tool for the prices between petrol stations.

Finally, the CPC is also seeking to update its forensic toolkit (software and hardware), which is likely to include computational tools as well. The Commission is aware that the purchase of contemporary software products is important for its work, however, it also acknowledges that the technical preparation of its staff is even more vital. For this reason, the CPC is actively enquiring other NCAs for the practices and experience with the toolkits available to them.

Canada
Canada's Competition Bureau

The Digital Enforcement and Intelligence Branch, otherwise referred to as 'CANARI' (Competition through Analytics, Research, and Intelligence), is Canada's Competition Bureau ('Bureau') newest branch. It strengthens the Bureau's work at every stage of investigation and acts as a center of expertise for:

- Intelligence collection and analysis;
- Behavioral insights;
- Remedies and monitoring;
- Data science and analytics;
- Digital tools for investigators;
- Technology insights, and
- Design thinking to support ideation and innovation.

CANARI's objective is to provide support and expert advice for colleagues across the Bureau in order to make the organization's work more efficient and effective. The Data and Analytics Team uses advanced analytics and data science to build tools that create efficiencies and add new capabilities to our investigative and promotional work. This includes projects with a focus on automation. The Technology Insights Team focuses on understanding evolving business practices and technologies, along with their impact on competition. The team helps identify emerging threats to competition more quickly and provides expertise to on new technologies.

More specifically, some examples of work already underway at the Bureau, include the following.

I. Intelligence

- The Bureau's Intelligence team works together with our digital teams to collect and analyze data and uncover trends that can be identified to support proactive enforcement. They also work together using statistical techniques and computational methods to proactively detect anti-competitive bidding practices in Canadian markets.

II. NLP Work

- Use of comparison tools to identify the differences between similar documents (within Nuix Discover);
- Predictive coding for document review in both investigation and litigation phases of investigations (within Nuix Discover);
- Concept clustering to group documents with common themes and identify connections between the cluster groups in order 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);
- Using Natural Language Processing (NLP) techniques with other techniques to automatically identify and measure groups of similar documents for investigations.

Catalonia
Catalan Competition Authority

For over a year, the Catalan Competition Authority ('ACCO') has been developing a computer tool named ERICCA, which uses big data and artificial intelligence to assist economists and lawyers working in this organization.[6] Particularly, this assistance includes both, simplified access to public procurement data in the context of bid rigging investigations arising from complaints and inquiries made to this organization, as well as a tool that facilitates the proactive identification of public tenders where it is most likely that participating companies have established collusive agreements for the submission of their proposals.

To access information related to Catalan public tenders, in mid-2022 ACCO signed a collaboration agreement with the Department of Economy and Finance of the Generalitat de Catalunya, the data owner of the Public Contracts Register (RPC).[7] The RPC is the electronic register containing the basic data of the public contracts awarded by the contracting bodies of the Catalan government and the Local governments, as well as the rest of entities subject to the public sector contracts law.

This agreement opened access to a broader range of tender-specific information, since information gathered from RPC's website before the agreement was less complete.[8] It made it possible to access details from many tenders, including not just the name of the winning company, but also the names of all other participating companies and their respective bids. This extensive information is vital for conducting analyses that aim to identify tenders where participants are suspected of having had a coordinated behavior.

Currently, the tool has already been able to identify several groups of companies that, in certain tenders, could have reached anti-competitive agreements, which would have allowed them to win the tenders without incurring in competitive

[6] The Catalan Competition Authority (ACCO) is an organization whose mission is to ensure the correct competitive functioning of markets in all productive sectors of the economy in Catalonia. Autoritat Catalana de la Competència, https://acco.gencat.cat/ca/l_acco/que_es_i_que_fa_l_acco/index.html (last accessed June 7, 2023).

[7] Contractació pública, https://contractacio.gencat.cat/ca/contractacio-electronica/registres-electronics/rpc/ (last accessed June 7, 2023).

[8] Previously, information was obtained using web scraping techniques from the Public Procurement Services Platform. Contractació pública, https://contractaciopublica.cat/ca/inici (last accessed June 7, 2023).

discounts, thus potentially obtaining supra-competitive profits. These tenders and companies involved in them are being analyzed thoroughly by ACCO.

However, we have observed that, despite the large number of tenders collected in the currently used database, about 250,000, corresponding to tenders carried out between 2010 and 2023, it is a relatively small sample compared to the total amount of tenders carried out during this period.[9]

This limitation on the data representatively encouraged us to implement our first improvement in the tool. We decided to complement our current database with data from the OPEN DATA portal, which provides information on nearly 2.000,000 tenders, significantly expanding our data resources.[10] The OPEN DATA portal includes sets of data produced or collected by public bodies that public administrations make available to citizens so that they can use them freely in a simple and convenient way.[11] Although this database does not include the participating companies in each tender, the name of the winning company and the amount of the award are indicated, information that can be very useful, among other things, to be more efficient in identifying collusive rotation agreements (e.g. possible temporary distributions or tendering body distributions).

Secondly, we have continued to advance in the incorporation of additional statistical parameters, identified in the specialized literature, which should facilitate the detection of anti-competitive behaviors. Currently, for each group of companies that the artificial intelligence algorithm clusters, the software:

- Calculates the individual discount of each company and the group as a whole and shows the average coefficient of variation of these variables (statistic that measures dispersion).
- Indicates the number of times that companies in the same group have made exactly the same economic offer (*i.e.,* suspicious circumstance of a possible agreement).
- It also calculates the percentage of tenders won by a specific company, taking into account the name of companies participating in the tenders.

[9] Considering each lot of the same tender as an independent tender.
[10] Dades Obertes Catalunya, https://analisi.transparenciacatalunya.cat/Sector-P-blic/Contractaci-de-Catalunya/hb6v-jcbf (last accessed June 7, 2023).
[11] Dades Obertes Catalunya, https://governobert.gencat.cat/ca/dades_obertes/dades-obertes/que-son-dades-obertes/ (last accessed June 7, 2023).

> This helps identify potential 'accompanying offers' that may have been designed to increase the likelihood of a particular company winning.

Another criterion for spotting anti-competitive behavior that we are adding to the algorithm involves segmenting by economic sector (CPV) and geographic location.[12] This will help us identify potential agreements to divide territories among companies.

Thirdly, we are investing in enhancing the usability and designing a user-friendly interface. For instance, the tool currently can display all tenders from the same group of companies, which can be sorted by various criteria, such as the date of award or the tendering body, allowing users to visually check for any rotation among the winners. To make this detection easier, we are introducing indices (with values ranging from -1 to 1) that calculate whether there's a rotation in the awards among the group's companies.

Moreover, the tool is being improved in order to allow users to easily download the information consulted in CSV format, which can be included in the working documents and reports or can be imported to other computer programs to complement and enrich the analysis with other software.

Lastly, it is worth mentioning that several institutions and public organizations have reached out to ACCO, interested in understanding how the tool was developed and in exploring potential collaborations.

[12] The CPV nomenclature (Common Procurement Vocabulary) is a system of identification and categorization of all economic activities that can be contracted through tendering or public competition in the European Union.

Chile
Chilean Competition Authority

I. Using a graph database to efficiently ensure compliance

The Compliance Enforcement Division of the Fiscalía Nacional Económica ('FNE') is responsible for the enforcement of mitigation measures established in concentration proceedings, compliance with resolutions of the Competition Tribunal or the Supreme Court, the prosecution of other infringements related to interlocking directorates and gun-jumping, and compliance with the minority shareholding notification obligation, contemplated in Article 4 bis of Decree Law Nr. 211. In particular, the Division has developed substantial experience with computational tools for the detection of infringements to Article 4 bis. The article establishes a legal obligation to inform the FNE of any acquisition of shares or capital in competitors that meets the following requirements:

- The acquisition exceeds 10% of the shares or capital of the competing company.
- Both the acquirer and the acquiree, or indistinctly each business group, have exceeded 100,000 UTM (or USD 7MM app.) in revenues during the last calendar year.

The enforcement of Article 4 bis has been done through a platform specially created to detect minority shareholdings that fall above the preceding thresholds. This platform uses information from several state bodies such as the Chilean Income Tax Service and the Financial Market Commission. The data gathered is processed using *Neo4j*, an open-source graph database.[13]

Neo4j allows us to build conglomerate structure charts that show economic agents linked by their ownership relations. The application can perform specific and general queries, such as the display of the nearest neighbors, identifying a particular identification number of a company, or filtering all the minority shareholdings that have a mandatory obligation under Article 4 bis. To illustrate the potential of this tool, *Figure 1*—top—shows the corporate group of an economic agent present in several sectors of the economy, while the bottom displays a zoom-in of the nodes.

[13] Neo4j, *Neo Technology*, https://neo4j.com/ (last accessed June 7, 2023).

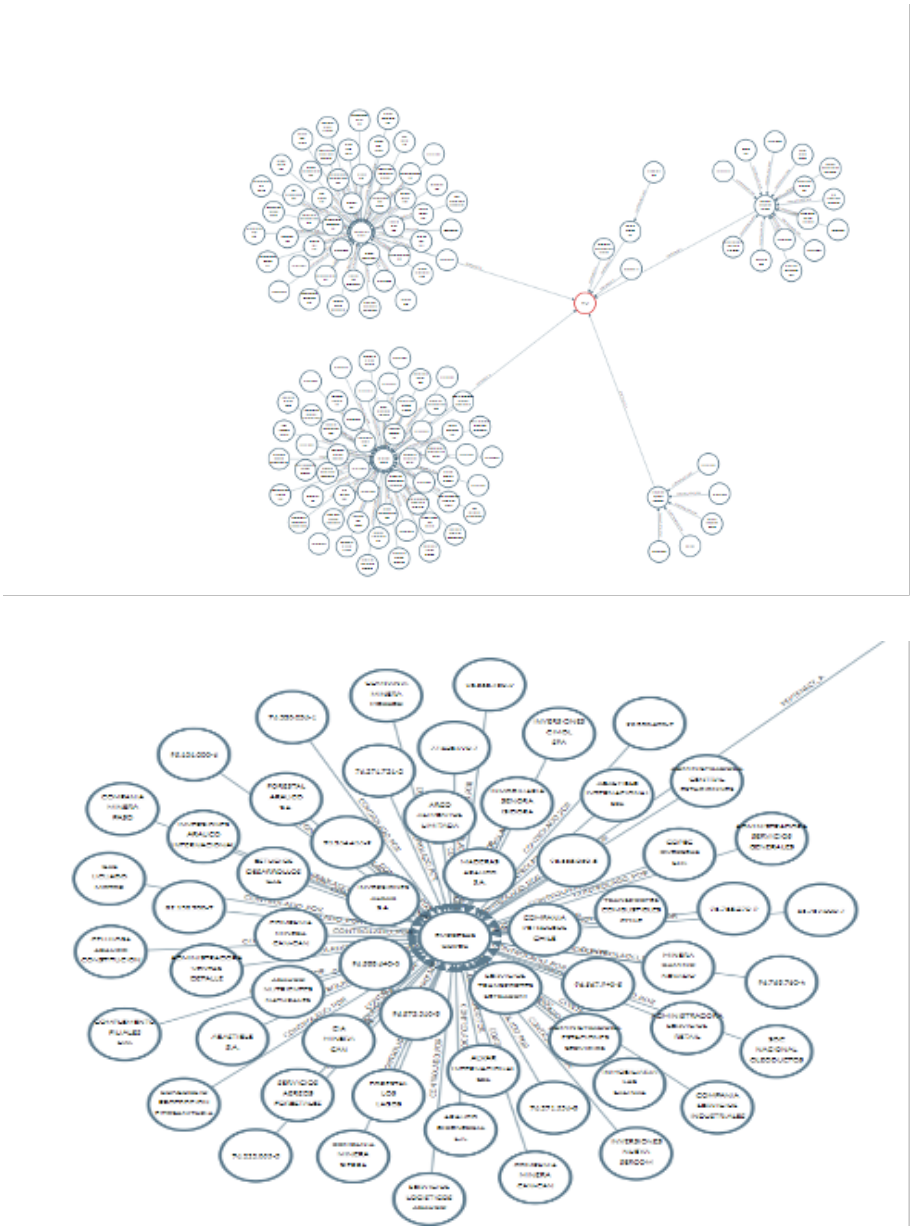


Figure I: Visualization of ownership relations.
Top: Corporate group.
Bottom: Zoom-in.

In this representation, which technically corresponds to a *directed graph*, each node corresponds to a natural or juridical person while the link between two nodes represents a shareholder relation. The direction of the arrow indicates that the agent at the *tail* (or origin) owns shares of the agent at the *head* (or destination). When clicking over a node, a list of information is displayed, including the economic activity of the respective person and its annual sales.

The tool allows us to filter all shareholdings between competing entities that met the conditions stated above for Article 4 bis and were classified under the same economic sector obtaining a list of potential cases to analyze. This procedure is straightforward and does not require much coding, as can be seen in the code snippet in *Figure II*. The result is a time saving application that contributes to a more efficient detection, considering that the alternative would have been a time and resource-consuming approach based on case-by-case examinations, or on eventual formal denouncements.

Notwithstanding, the list must be further refined by human analysis because of potential errors in the database and because economic activities listed in the database are too broad. Therefore, the economic activities must be further specified using mainly public information, to determine that there are in fact rival products or services between the groups linked by the shareholding and, therefore, an infringement.

```
MATCH (g0:Grupo)<-[:PERTENECE_A]-(gc0:Controlador)<-[:CONTROLADO_POR]-(c0:Contribuyente)-
[r:SOCIO_DE]->(c1:Contribuyente)-[:CONTROLADO_POR]->(gc1:Controlador)-[:PERTENECE_A]->(g1:Grupo),
        (g0)<-[:ACTIVIDAD_DE]-(a:Actividad)-[:ACTIVIDAD_DE]->(c1)
WHERE g0.venta_min_UF > 100000
AND c1.venta_min_UF > 100000
AND r.participacion >= 10
AND g0 <> g1
WITH g0, gc0, c0, r, c1, gc1, g1, a
MATCH (a)-[:ACTIVIDAD_DE]->(d:Contribuyente)-[:CONTROLADO_POR]->()-[:PERTENECE_A]->(g0)
RETURN g0.nombre as `Grupo declarante`,
c0 as declarante,
r.participacion as Participacion,
c1 as adquirido,
g1.nombre as `Grupo adquirido`,
a.act_econ as `Act. econ.`,
d as competidor
```

Figure II: Shareholdings filtering query example

In 2020, after a month of implementation, the FNE filed the first two lawsuits supported by this tool against Inversiones Los Orientales[14] and Banmedica,[15] for the minority interest they acquired in competitors without informing the FNE. In both cases the FNE agreed to an extrajudicial settlement where the companies involved committed to pay a substantial fine.

[14] FNE, *Lawsuit against Inversiones Los Orientales*, https://www.fne.gob.cl/wp-content/uploads/2020/01/Requerimiento-FNE-Inv-Los-Orientales.pdf (last accessed June 7, 2023).

[15] FNE, *Lawsuit against Banmédica*, https://www.fne.gob.cl/wp-content/uploads/2020/01/Requerimiento-FNE-Banmedica.pdf (last accessed June 7, 2023).

Still, there are some challenges ahead such as getting more precise data of the economic activities carried out by companies and their annual sales and accessing up-to-date data on the ownership relations each time a query is made. Another medium-term objective is to extend the capabilities of the application to detect companies engaging in gun-jumping behavior.

II. Creation of the Intelligence Unit

The Intelligence Unit ('IU') of the Fiscalía Nacional Económica was launched in September 2020, and has a staff of three professionals, including two data scientists. Through technological tools and analytical techniques, the IU focuses on cartel screening and the development of investigative approaches for the analysis of large volumes of data, structured and unstructured.

The FNE has prioritized an in-house development strategy to embed, in these tools and techniques, the specific knowledge of lawyers and economists working in the office, and, by doing so, fulfilling in a better manner the investigative needs of the institution.

Below we present the IU's main ongoing projects.

A. Cartel detection

- Automated web-based data collection tools to collect information that could help answer consumer complaints, improve the data used in ongoing investigations, and build ex-officio investigations. We are currently developing the implementation of *Machine Learning* algorithms and other *Artificial Intelligence* approaches to help us analyze and classify the processed information with a focus on detection.
- Development of a platform to analyze public tenders based on screening techniques, which requires sophisticated data collection and processing algorithms to display simple and useful reports in a user-friendly interface, reducing the time required to gather and review large quantities of data and identify anticompetitive behavior in this kind of procurement mechanism. In a future stage of this project, Machine Learning algorithms could be implemented to build classification models that use the available historical data to help assess the existence of bid-rigging practices in public tenders and build ex-officio investigations.

B. Investigative systems for data analysis

- Systematization of data provided by public services, such as the Customs National Service. In a future stage of this project, this data should be uploaded on a platform that can make it accessible for all the divisions of the FNE to use it as they see fit to perform the analysis required to accomplish their tasks. Public and private alliances are key to develop search engines that give the staff useful data for their investigations, reducing considerably the timed needed to find relevant information.
- Systematization and analysis of phone calls logs and georeferenced data obtained from mobile devices, using programming tools to manage large datasets and create valuable reports to help ongoing investigations by identifying useful trends to improve the decision-making processes and the storytelling of the cases.

Colombia
Superintendence of Industry and Commerce

As reported in 2021, the Superintendence of Industry and Commerce (SIC) 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. In addition, in 2022, the SIC has begun to develop a new data analytics tool named 'Búho.'

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.

The current results can be summarized as follows.

I. Sabueso

Sabueso is a price monitoring data tool aimed at enhancing the authority's inspection, surveillance, and control functions. It has the objective of harnessing available data on products offered by supermarkets on their websites. This tool allows 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.

During 2021 and 2022, Sabueso captured information on one (1) store in different categories. For 2023, the coverage will be increased to five (5) market stores, and the tool will be programmed: (i) to detect possible anticompetitive conducts between the monitored agents, and (ii) to send alerts directly to the authority.

II. Inspector

This data analytics tool developed by the SIC supports the monitoring process by the Competition Advocacy Group of different regulatory projects published on the websites of state regulators. This tool alerts the authority of published regulatory projects detected on their official websites that may impact competition.

The tool has a model for identifying regulation projects that could affect the competition dynamics in the market. It performs natural language processing (NLP) and converts the document into a vector that will go through a classification model to determine the possible impact of the project on competition matters. The model has been trained from the history of concepts issued by the Competition Advocacy Group.

In July 2022, Inspector was monitoring and reporting relevant information of 38 out of 68 different Colombian authorities. In comparison, in April 2023 Inspector was monitoring 60 out of the 68. This means, Inspector increased its regulatory projects issued capability detection from 55.8% to 87.1% the monitored authorities.

In conclusion, Inspector has evolved in a highlighted way due to its effectiveness in gathering information on new regulatory projects issued by regulatory entities to analyze later whether such projects may impact free competition.

III. Sherlock

Sherlock is the data analytics tool project that seeks to support the SIC's investigators in the identification of signs or patterns that suggest potential anticompetitive behaviors on public procurement processes (bid-rigging).

This data analytic tool is currently in its second stage. This means that the Data Analytics Group is working on automation of the search of the abovementioned signs and patterns within the bulk of public procurement data available online based on machine learning technologies and deep machine learning. It relies on artificial intelligence (AI).

This tool hasn't been used to identify possible bid-rigging conducts because it is still in a developing stage.

IV. Búho

This data analytics tool is meant to support the SIC in the detection of different news or reports circulating online, on different websites, that could be potentially related to anticompetitive behaviors. This way, the authority gets the alert and can evaluate the scope of the reported facts.

Búho is still in an early developing stage. Currently, the Data Analytics Group of the SIC is identifying: (i) which websites are going to be the source of information; (ii) key words to generate the alert; and (iii) the best way to warrant the effectiveness of the notification to each related group.

Cyprus
Commission for the Protection of Competition

The eOASIS system was developed by the government through co-financing with the European Union.[16] The eOASIS office automation system is the Government's Official Correspondence and Communication Management System both within the Service itself and between the Services of the various departments of the government. It concerns the electronic filing and management of official documents of government organizations, as well as the automation of the procedures and regulations governing the creation, archiving, security, circulation of documents as well as their final destruction or their long-term preservation and future availability to the public or to researchers through the State Archives. The eOASIS System aims to automate the management of documents (Electronic Document Management) and is web-based in accordance with European and other international standards.

It covers all the requirements of the public sector in relation to document management by providing a comprehensive range of functions. The system supports, among other things, the creation of electronic workflows, which automate predetermined procedures, such as that of filing, and allows remote working, thus serving the operators who need access to documents of their service when they are out of the office.

The online version of e-OASIS has a central infrastructure (eOASIS Cloud) from where it serves all the Government Organizations in which the system currently operates, with the exception of three independent facilities.

By using the eOASIS system, the Services benefit from the following:

- A significant reduction in the time of filing and handling documents;
- Increased document security and controlled access to documents by authorized users;
- Easy and fast retrieval of documents based on search criteria;
- Simultaneous access to the same document/folder by multiple operators;
- Keeping all documents in electronic form and significantly reducing the use of paper;

[16] The initials of the system stand for Electronic Office Automation System and Integrated Services.

- Access to the system from remote points and possibility of teleworking from abroad and/or home;
- Maintaining the integrity of the documents (avoid their alteration and/or falsification);
- Avoiding the loss and/or destruction (damage) of physical folders and documents;
- Automatic electronic transfer of important and/or historical documents from an organization to the State Archives for permanent safekeeping and long-term preservation;
- Improvement of communication and cooperation between departments;
- Reduction of the time of sharing documents between departments;
- Increase of productivity, efficiency and effectiveness of the government;
- Containment of the growth rate of specific staff groups, such as secretarial staff, clerks, auxiliary staff (callers), etc.;
- Reduction in the use of the post office, handling of physical envelopes through couriers, storage areas, unnecessary use of paper, faxes, etc.

The most important advantages of the System are:

- The use of a Unified Thematic Classification System, which provides uniform numbering of files throughout the public sector;
- The possibility of simultaneous access to the same document by more than one person;
- The ability to find documents quickly and easily, through various criteria (eg title, date, file number, etc.) as well as through content search through optical character recognition (OCR);
- The ability to access documents from remote locations;
- Ensuring controlled access to the system and increased document security;
- The reduction of paper use;
- Increasing productivity through the immediate availability of documents, information and instructions;
- The reduction of document filing time.

The System, at the present stage, operates in each department/service independently, however it provides the possibility of inter-departmental connection (*i.e.,* electronic sending of documents between Departments/Services through the System) which is piloted in 4 Departments (Deputy Ministry of Research, Innovation and Digital Policy, IT Services Department, Public Administration and Personnel Department).

The System is currently in operation in 49 Organizations (Ministries/Deputy Ministries/ Departments/Services) of the public sector. The implementation of the eOASIS System is expected to be completed throughout the public ssector by the end of 2026. It is noted that the Office Automation System, eOASIS is operational in various organizations of the wider Public Sector, such as the University of Cyprus, the Research Promotion Foundation and the Cyprus Energy Regulatory Authority.

The system was introduced for the electronic management and filing of documents, and it is further developed by the central government. It is a system applied in the public service of the Republic of Cyprus and it was introduced in our offices in December 2020. The CPC notified all interested parties regarding the introduction of this system through a press release calling for the filing of documents in a digital format.[17]

More information (in Greek) can be found on the Commission's webpage.[18]

[17] Commission for the Protection of Competition, http://www.competition.gov.cy/competition/competition.nsf/All/26B46D540F515708C2258634004545FE?OpenDocument&highlight=e-oasis (last accesed June 7, 2023).

[18] Commission for the Protection of Competition, https://dits.dmrid.gov.cy/dmrid/DITS/dits.nsf/All/46D813D43EFFDA97C2258565003A4B73?OpenDocument (last accessed June 7, 2023).

Czechia
Office for the Protection of Competition

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

I. Project Watson

The Office is still testing this project whose initial goal is to gather all the information sources relied on by our employees in a single platform. This tool should be able to search and find relevant outputs in many languages and filter non-structured data with the help of Artificial Intelligence (AI); therefore, it should be able to evaluate relevance in an advanced way. Our employees will then have greater comfort in terms of efficiency and speed in searching through information sources. This project is not profitable; it only aims to secure more efficient search through these sources of information, but, at the same time, some pages which have been searched through are subsequently scanned by a robot and indexed into our program under certain keywords.

II. Project DATACROS II

The Office still participates in the DATACROS II project. In the case of this project, we see the potential of pointing out the risks of prohibited agreements in certain areas. We will be able to trace the actual owners of companies and obtain relevant information for, especially, reveal ownership structures. With this project, the Office could be more successful in bid-rigging detection.

Individual activities associated with the project are still ongoing. Two workshops were held in June and July 2022 to present the main functionalities of the upcoming tool and its resources and to define the data and specifications of functionalities for the first test approach of the proposed tool. In March 2023, a meeting on DATACROS II was held to present the tool in detail. The Office has started testing its functioning and will subsequently interact with the other competition authorities involved in the project.

III. Strategy for Electronic Public Procurement for the Period 2021-2025, Cooperation with the Ministry of Regional Development

The Office remains part of this strategy in 2023, with its key activities mainly consisting of methodological contributions and raising awareness, and the provision of know-how in the field of public procurement and competition protection.

IV. The project of Masaryk University Brno (Czech Republic): New Mobility--High-Speed Transport Systems and Transport-Related Human Behavior

The objective of the project is to intensify intersectoral and interdisciplinary cooperation in the framework of research into changes in the mobility behavior of the population within the context of the development of high-speed railways in the Czech Republic. In the framework of the project, a specialized platform will be created, interconnecting the research institutions, entities active within the relevant domains, and public bodies, which will comprehensively address the socioeconomic impact of this phenomenon on society and the overall quality of life. The purpose of this platform is the effective exchange of experience and knowledge between individual participants.

V. Completed Projects within the Technology Agency of the Czech Republic under the Auspices of the Office's Competition Division

The methodology for the supervision of public tenders in passenger railway transport and detection of infringements of competition rules was developed in 2015 as part of the grant project Optimization and Improvement of Competitive Tendering in the Passenger Railway Transport in the Czech Republic according to EU Directives. This project aimed towards creating certified manuals for public policy making in the area of competitive tenders in passenger railway transport. The main users of these manuals will be the Ministry of Transport and the competition authority. Such manuals should provide a coherent basis for competitive tendering designs, making them effective and minimizing any potential conflicts between the agendas of these institutions. The methodology contains a description of basic pricing methods to detect bid rigging.

Automatic detection of suspected collusive behavior in public procurement was developed in 2017 as part of the Omega project. The project resulted in the detector

and *Probber* software, including relevant methodologies and manuals for detection of bid rigging.[19]

The Decision support system for merger assessment in markets with homogeneous product and spatial differentiation was developed in 2020. Within the project, the *MergerSim* software was developed to assess mergers in the fuel market.

VI. Future plans

In the past, we developed a web scraping tool called *Probber* in cooperation with Masaryk University. This allowed us to download data from the public procurement bulletin (a publicly available search engine for public procurement forms) more easily and create an internal database. The Office wants to extend the database by adding additional data, especially from the profiles of contract authorities. We would like to consider other sources, *e.g., ROZZA*--tender signpost. We also want to build new tests/tools for bid-rigging detection (implement new tests working with competitors' bids, experimenting with Machine Learning Algorithms--*Random Forests*, *Lasso*, and *Ridge regression*, etc.).

[19] *Probber* stands for *procurement + robber*.

Finland
Finnish Competition and Consumer Authority

The Finnish Competition and Consumer Authority (FCCA) has recently introduced a cartel screening function that is partially based on automated computational tools. In the process, applicable data is screened using tried and tested statistical methods. The screening consists of two stages where suspicious markets are first initially flagged and then reported internally to the cartel detection function when considerable collusive tendencies are observed. The analytical work is conducted by the agency's economists.

For the time being, the FCCA focuses on discovering cartels in public procurement tenders. The agency has good access to comprehensive bidding information from tenders held in Finland which makes screening this data a good starting point. However, the scope of cartel screening can be broadened to cover data from other types of markets e.g., by implementing the tools on data collected through web scraping.

To showcase the impact of cartel screening, the FCCA applied the toolkit on two substantial Nordic cartels which operated in the asphalt paving markets in Finland and Sweden. Cartel behavior could be distinguished from the distribution of bids both using a supervised machine learning approach and a separate distributional regression test. The results clearly demonstrate that the cartels could have been caught using statistical methods. With these new ex-officio cartel investigation tools, the harmful conducts could also likely have been detected earlier.

The FCCA's cartel screening process consists of two interrelated stages. In the first one, a set of methods are applied systematically on all different combinations of tenders that might need to be looked at together. This part of the process is highly automated in order to achieve maximum efficiency. In practice a multitude of different generalizable tests with different types of thresholds are ran simultaneously, and consequently some of the markets are flagged as being initially suspicious.

In the second stage of the process, a select group of markets are analyzed in-depth using methods that require more customization. The aim of the second part is to weed out truly suspicious firms and markets from false red flags. The guiding thought is to test whether the results from the first stage could be explained by

something other than collusion. The second stage relies less on automation as the underlying data needs to be meticulously examined. Markets and firms still deemed suspicious after the second stage are reported internally to the cartel detection function which decides upon the next steps of the investigations.

France
Competition Authority

In 2020, the French Competition Authority ('Autorité') decided to create a dedicated digital unit in order to strengthen its resources in digital areas.[20] 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 Autorité is proud to be part of the Stanford Computational Antitrust and to have contributed to its first annual report. The following presents what has been done since then.

The Autorité, in collaboration with CodeX Computation Antitrust, has published on its website the first interactive network graph tool capable of identifying within the Autorité's publications (*i.e.*, the decisions, opinions and interim measures published between 2009 and 2021) the references to its other publications and to represent these interconnections in a graph.[21]

Designed for the antitrust community (case handlers, scholars, lawyers, etc.), it takes the form of a network graph in which the Autorité's publications are represented and connected with each other by their citations. It allows the identification of the interconnections between different publications at a first glance and gives an overview of the Autorité's antitrust cases. It also can be used to:

- Get a visual overview of French antitrust cases;
- Create a bibliography for research projects (starting with a reference or a sector, the user will be able to use the tool to fill in the gaps and find closely related publications;
- Identify all key publications: users can visualize important French antitrust publications based on their impact on newer publications (number of citations) and the importance of web audience;
- Discover the most relevant prior and derivative works.

[20] Autorité de la concurrence, *The Autorité Creates a Digital Economy Unit*, https://perma.cc/6742-UKUX (last acccessed June 7, 2023).
[21] Digital Economy Unit & CodeX: Stable version 1, https://sen-codex.dev/ (last accesed June 7, 2023).

An article explains all the process and the visualization tool developed by the Autorité.[22] Moreover, the entirety of the data are accessible as open data on the Autorité's GitHub.[23]

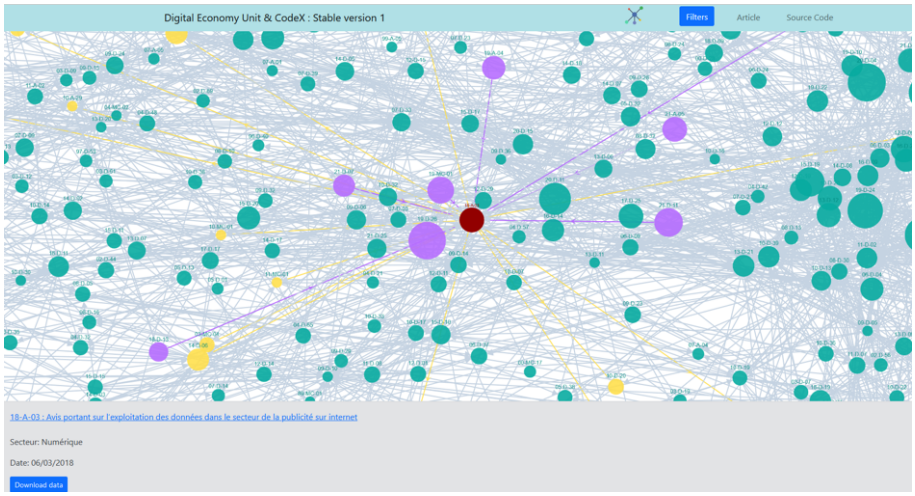


Figure I: Highlight of opinion 18-A-03 in the tool and its related publications[24]

The Autorité is also still 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).[25] It will be useful for collaboration with other agencies in the field of cartel detection. 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 built).

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

[22] Yann Guthmann, Adrien Frumence & Camille Hoogterp, *Deploying Network Analysis in Antitrust Law*, https://www.autoritedelaconcurrence.fr/sites/default/files/2023-01/Stanford-Computational-Antitrust-en.pdf (last accessed June 7, 2023).
[23] Autorité de la concurrence, *CodeX Networkgraph*, https://github.com/AutoriteDeLaConcurrence/publication_sen-codex_networkgraph (last accesed June 7, 2023).
[24] Opinion 18-A-03 of March 06, 2018 (*Data Processing in the Online Advertising Sector*).
[25] Transcrime, *DATACROS II Kick-off Meeting*, https://www.transcrime.it/en/datacros-ii-kick-off-meeting/ last accesed June 7, 2023).

Greece
Hellenic Competition Commission

I. Introduction

The Hellenic Competition Commission ('HCC') has been one of the first movers, among competition authorities in Europe, in the study and integration of advanced computational competition law and economics tools in its day-to-day activities.[26] Early on taking over the Organization, in 2020, the new Agency leadership realized the importance of AI and data science tools in enhancing the pace and scope of competition law investigations and put in place a holistic strategic vision with the aim to transform the HCC to an AI-augmented competition authority by 2024.

In pursuing this overall ambition, the HCC first completed in 2020 a comprehensive inception report with the aim to map the use of different computational techniques (Big Data, AI, machine learning, deep learning) by competition authorities around the world, focusing in particular on some selected competition authorities, in view of their different level of development, size and legal system in order to better understand the linkage between the use of these computational techniques and institutional change (HCC, Inception Report, Computational Competition Law and Economics, 2020).[27] In the meantime, the HCC invested in the construction of a bespoke data infrastructure for the collection and analysis of data and economic intelligence, as well as proceeded to the procurement of number of commercial off-the-shelf AI software products (*e.g.*, Tovek Intelligence) for use by the newly formed forensic and market mapping units of the HCC (2020).

The HCC Data Analytics and Economic Intelligence Platform (DAECI) was conceived of as a tool to integrate and keep updated multiple external data sources in common database schema with the aim to provide visualization tools for data exploration and a screening device to HCC's economists and lawyers. The platform is being hosted within the premises of the HCC to the users having the appropriate access credentials. The application is being deployed as a Flask Server that incorporates a number of dashboards that enable various forms of data visualization

[26] Written by Ioannis Lianos, Vassilis Vassalos, Petros Boulieris, Maria Niki Fourka, Nikos Delis, Xanthos Xanthopoulos, Lampis Tzai.

[27] Hellenic Competition Commission, *Computational Competition Law and Economics Inception Report*, https://www.epant.gr/en/enimerosi/publications/research-publications/item/1414-computational-competition-law-and-economics-inception-report.html (last accessed June 7, 2023).

(e.g., box plots, basket plots, data error bars). The Platform also integrates a screening method to detect anti-competitive practices--including cartels and excessive pricing- from the analysis of market data (in particular prices, but also for some products quantities), 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 2200 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 interactive dashboards are created by utilizing the framework of *Plotly-Dash* that is known for its usability features and scalability.[28] The screening tool was implemented in number of investigations opened by the Hellenic Competition Commission since 2020. Moreover, the Commission is investing in its expandable Big Data Management Infrastructure Platform/ dashboard, for the collection of real-time public data from different sources (Price Observatory of Supermarkets, fuel prices, vegetables and fruits prices, public procurement data, etc.), data that is automatically uploaded and updated every day or many times per week. Furthermore, the HCC has recently (2023) proceeded to cooperation with experts to design a program, drawing raw data from unstructured information available in the national public procurement database and other sources. This data will be mainly used for cartel-detection but will also offer an integrated data analytics environment with various tools/apps, on the basis of bespoke programs and /or available off the shelf software tools to visualize and analyze data.

In order to successfully complete the above-mentioned holistic strategy, the HCC completed in April 2021 the recruitment process of a chief technology officer and a team of data-scientists who joined the HCC for a renewable mandate of two years. The team is headed by Professor Vassilis Vassalos, professor of informatics at the Athens Economic University who is the Chief Data Scientist of the HCC, three data scientists (Petros Boulieris, Maria Niki Fourka, Lampis Tzai) and two computer scientists (Nikos Delis, Xanthos Xanthopoulos). The team has been working on several programs requested by HCC's management with the aim to transforming all aspects of the work of the HCC. The projects undertaken relate (i) to the development of data analysis tools to facilitate HCC's staff in mapping and monitoring daily markets, (ii) the construction of bespoke algorithms for the purposes of HCC's investigations and more generally the implementation of machine learning and data analysis models to reduce the time required to analyze cases, since HCC employees will be able to take on and process more cases at the same time, (iii) the development of reusable code to

[28] Plotly, Dash Enterprise, https://plotly.com/dash/ (last accessed June 7, 2023).

accelerate common workloads, (iv) the establishment of a project management platform so as to enable the day-to-day monitoring of the work of the HCC's employees with the aim to enhance more flexible forms of working (teleworking) and administrative efficiency.

This brief note will focus on some of the computational competition law and economics applications of the HCC in different work streams. In particular, the Data Science team at HCC has contributed in (i) mapping markets and harvesting data from supermarkets/groceries through the development of the DAECI platform, (ii) providing data analysis for the ongoing fuels market investigation, (iii) assisting the HCC in the monitoring of the household basket data tool put in place by the Ministry of Development, (iv) engaging in the technology-assisted review of data in the inter-bank collusion investigation.

Furthermore, we will succinctly present some of the ongoing workstreams of the data science team, in particular (i) the implementation of an AI-enabled document library to enable semantically similar file retrieval, (ii) developing tools for social media and company monitoring, and (iii) proceeding to an AI-revamp of the HCC's existing whistle-blower platforms.

II. Existing workstreams in the work of the data science team

A. The HCC Data Analysis and Economic Intelligence Platform ('DAECI')

1. The construction of the DAECI

When conducting legal investigations, the Electronic Discovery Reference Model is a comprehensive approach to digital data. The EDRM has 9 basic stages: Information Governance, Identification, Preservation, Collection, Processing, Review, Analysis, Production and Presentation. The HCC is increasingly faced with digital data when conducting legal investigations, requiring a significant investment of time in data collection, processing, and analysis, so streamlining more common and routine tasks is important.

The purpose of the HCC Data Analytics and Economic Intelligence (DAECI) platform is to create an easy-to-use data analysis environment, with the aim of collecting market data from publicly available databases in real time, storing them in a local database, editing, visualizing and analyzing them, as well as the development

of screening tools using Machine Learning in order to detect anti-competitive practices (market screens) in various sectors of the Greek market.

The main sources for data collection so far (in the first year of operation of the DAECI platform) are e-katanalotis (Market Observatory) for monitoring food prices as well as the basic products of the Greek household, OKAA (Central Markets and Fisheries Organization) and Eurostat for monitoring fruit prices, vegetables, meats and fish, as well as fuelprices.gr for fuel prices in the Greek market (*See* Table 1).

	Supermarkets	**Fuels**	**Fruits and Vegetables**	**Fish**
Brand/Sale Region	16 Brands	20 Brands	Wholesale	10 regions
Categories	123	-	3	5
Products	>1200	105	350	148
Years	2014 - 2016 2020 - 2023	2015 - 2018 2020 - 2023	2017 - 2023	2018 - 2023
Data Frequency	Daily	Daily	1-3 times per week	Monthly
Type of Prices	Shelf price	Offering Prices	Offering Prices	Min-Mean-Max of selling price
Quantity		-	-	Yes

Table I: Price monitoring

The platform uses existing APIs to extract the information it needs, but possesses the extensibility to manage other types of data that can help the work of the Competition Commission such as standalone files (*e.g.*, .csv, .xlsx, .pkl) and/or direct links to other databases within the Commission. After data collection there is always the possibility of errors during data entry, especially if done by a human. For this reason, input data are processed in order to remove anomalies such as gaps at various intervals, inconsistent data types (typographical errors, etc.) or incorrect categorization. The platform, to solve the above, undertakes the creation of new product categories and their integration into its database using Natural Language Processing (NLP).

The DAECI platform provides the user with specially designed dashboards for market monitoring so that quantitative analysis can be performed immediately, and reporting can be made easier. The user can select the type of visualization from various types of diagrams (box plot, time series diagram, deviation range plot, custom basket plot) and at the same time make changes to its parameters to make it easier to explore the data during modeling time. In more detail, the time series chart identifies price trends for different products and categories, for one or more companies. The graph gives a more detailed picture of price changes by category or product using measures such as quartiles, median, safety margin and extremes. The deviation range diagram describes the minimum, average and maximum selling price of catches per month, for each month.

To create the custom basket diagram, a default basket was created, indicative of the average Greek household, with 16 popular categories based on the analysis of the Consumer Goods Research Institute (IELKA), weighing the product categories with the 2019 Consumer Price Index Authority (ELSTAT). The above diagrams give the possibility to select different time units (daily, weekly, day of the week, monthly), selection of the appropriate meter per case (average, median), observation of values with or without seasonality, definition of the desired monitoring period, descriptive and statistics for selected quantities, data mining and direct comparisons between charts.

For the analysis of the data with Machine Learning algorithms, the experience and knowledge of HCC experts were used both for the design stage and for the implementation stage. So far, we have created a two-step screening mechanism to allow fast data-driven prioritization of Commission cases. Upon selection of a time period and some potentially suspicious products to investigate, the screening mechanism produces various relevant metrics (*e.g.*, T-tests, Log-Differences in

Differences) and a diagram to help the user assess whether the counterfactuals show unusual and suspicious changes in the prices of the products under consideration. Finally, the mechanism calculates the expected value of the suspicious product, and it is identified if there is a structural break in the time series. All of the above bring the Commission one step closer to its original goal which is none other than mapping the Greek market and having intelligent tools at its disposal to monitor the state of competition among firms.

2. Completing the DAECI with data harvested from e-shops

Furthermore, the HCC Data Science team has recently completed the DAECI with a more extensive and systematic data harvesting from e-shops with the aim to constitute a complete catalog of every product offered by supermarkets that operate in Greece and have an e-shop and the constitution of time series, starting in 2023. The data collected take the form of JSON files for every product 61300 products in total as of this writing) from every supermarket and includes the name of each product, the prices daily and the SKUs/Barcodes. The data also includes discounts, eventually promotion campaigns and offer stickers, when applicable.

The custom tools elaborated by the HCC Data Science team harvest automatically information from each e-shop daily and raw data is backed up before any preprocessing. Following this stage, the data is processed and inserted in the database. With this new tool the HCC stays up-to-date on the latest pricing & product information and can implement the structural breaks methodology that has been put in place for the DAECI platform at a larger scale, without also being dependent on access to the e-katanalotis app, which constitutes the main source of information on which relies the DECI. The project can be easily scaled up to include additional supermarkets. Other e-commerce platforms can be added in the future, and automating the data collection process would allow the HCC to save time & resources.

B. Providing data analytics assistance to the Fuels Market Investigation

Systematically using data analytics tools may largely enhance the capacity of the competition authority to complete more timely sector inquiries and market investigation references, in particular, if there are large data sets available.

One of the projects on which the HCC Data Science worked team in 2022 and 2023 concerned the preparation and processing of data for the Fuels markets mapping and Market Investigation Reference in 2022 and 2023. The team proceeded

to data collection about prices for the period Q3 2019 - Q3 2022 regarding different levels of the fuels supply chain:

- Price information for crude oil;
- Invoices from refineries to major wholesalers (Refiner → W);
- Invoices from wholesalers to gas stations (W → Retail);
- Store prices for gasoline, diesel, and heating oil at every gas station in Greece (Retail → Public).

While the estimated time to process data by hand was initially 2 months with the involvement of 4 employees, the HCC Data Science team as able to complete this task in 3 weeks, with the involvement of 2 employees. Data harvesting was complicated by the fact that not all information on prices was publicly available:

- There was price information for crude oil, although the team employed price imputation for days where the Brent & WTI Barrel prices were not available.
- With regard to invoices from refineries to major wholesalers (Refiner → W), most discounts were applied directly on the invoice, while others were granted through "Credit Invoices" that had negative values and were used to give discounts or bonuses to customers. The team also had to process the invoices to obtain the actual, final price per liter, taking into account discounts, credit invoices and VAT. The team matched the wholesaler names on the refiner's invoices to the wholesaler invoices using the wholesaler name as a unique identifier.
- With regard to invoices from wholesalers to gas stations (W → Retail), as before, the team had to process the invoices to obtain the actual, final price per liter, taking into account discounts, credit invoices, and VAT. It also matched the retailer names on the wholesaler's invoices to the retailer store prices using the retailer's name as a unique identifier. Note that many retailers had undergone a change of management, and the station name column changed as a result. This created a problem as multiple gas stations could share the same physical address, which would be impossible. The team used fuzzy string matching at first, then NLP to ensure that each station address uniquely described a station.
- With regard to store prices for gasoline, diesel, and heating oil at every gas station in Greece (Retail → Public), the Data Science team collected information at the station-service level from the public database fuelprices.gr which is station-reported and covers the whole country,

although it also proceeded to rectify some errors in the database as many vital pieces of it such as addresses, station names, and postal codes were incorrect or missing. The Google Maps geocoding API was employed to solve the issue.

C. Monitoring the 'Household Basket'

In the context of its anti-inflationary policies, the Greek Ministry of Development established the 'Household Basket Initiative' with the aim to increase competition among retails chains for categories of 'basic consumer goods' and therefore to limit the likelihood for sudden and steep price hikes. According to this initiative, food and other product hypermarkets that carry out retail activities (supermarkets), the total annual turnover of which exceeds ninety million (90,000,000) euros per year, based on the financial statements of the previous business year, are required to send to the Ministry at regular intervals, a list of consumer products ('household basket') that are necessary for a decent living of households and are available at affordable prices. The list of the products included in the 'household basket' for each party liable and their prices is set freely by the undertakings and products included in the 'household basket' are checked as a matter of priority by the competent supervisory authorities for compliance with the provisions on unfair profiteering. The products in the list may be differentiated each time a new list is sent to the Ministry of Development and Investments. Those liable list at least one product from each product category included in the household basket and ensure that it is available at an affordable price, compared to other products of the same category. They also ensure that the consumer, who comes to their physical stores or accesses their digital marketplaces, is informed about the composition of the 'household basket' that they propose by any appropriate means such as product catalogs, brochures and specific labeling on the products or at the marketplaces. The non-sending of a product catalog or the incomplete sending of a product catalog by the obligors shall incur a pecuniary administrative penalty of five thousand (5,000) euros for each day of delay in sending a complete catalog.

The HCC expressed the Opinion that the regulatory initiative does not raise concerns from a competition law perspective and, in any case, despite the distortions of competition that may be caused by this measure, it is proportionate, in view of the current difficult circumstances for consumers and given the temporary nature of the measure. However, the HCC also raised concerns with regard to the fact that that supermarkets selecting the products they wish to include in the household basket may offer them a leverage capacity in negotiations with suppliers of such products,

especially with those with a weak bargaining power, or that such initiative may lead to systematic self-preferencing practices by supermarkets for their own private label products, against branded products, while it cannot be excluded that there could be price increases on products or product categories (from the same or different suppliers) off-basket to compensate for any losses from making the in-basket products available at affordable (or possibly reduced) prices. The HCC also expressed its intention to monitor the implementation of the 'household basket.'

The HCC Data Science team was in charge of developing analytic tools that would facilitate the monitoring of the operation of the 'household basket' and controlling for anti-competitive behavior adopted by supermarkets, thus limiting the time and human resources necessary for performing this task. Thus, it put in place a useful and up-to-date dashboard that analyzes household basket data. Using this tool, the prices charged by the stores are monitored, enabling analysts to recognize abnormal trends in the pricing of goods. Clicking on each supermarket in the weekly basket total cost chart breaks down the basket structure for the selected supermarket. The basket structure chart shows the contribution of each category to the total price of the basket.

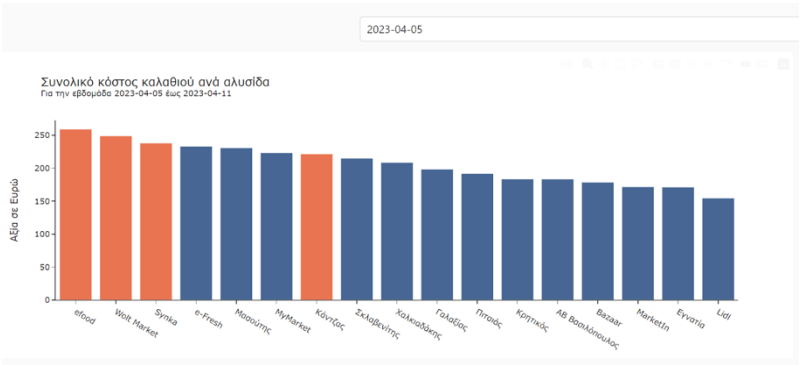


Figure I: Household basket

The HCC Data Science team also developed *between-category time series* by retailers, which shows the average price of product categories for each supermarket over the selected time period. Each supermarket is represented by a different colored line in the diagram. The price of each line represents the average price of products in that category in that supermarket. This chart can help analysts monitor product price trends by category in supermarkets and discover any price anomalies that may indicate anti-competitive behavior by supermarkets. Finally, the team also developed the *within-category time series by the retailer*, which shows the evolution of the price of all products belonging to a selected category for each supermarket. Each

line in the chart corresponds to a specific product and retailer and is colored differently for easier identification.

D. Technology-assisted review of data in the inter-bank collusion investigation

The HCC Data Science team has also worked on tools facilitating the technology-assisted review of data, in particular in the ongoing inter-bank collusion investigation, some of which are briefly examined below.

1. Attachment Analysis

In the Hellenic Competition Commission's (HCC) investigation, attachment analysis played a crucial role in filtering out irrelevant emails and identifying attachments that required further review.

To extract the text from each attachment, the data science team utilized Apache Tika, a powerful text extraction library. This enabled the extraction of text from various file formats, including .doc, .pdf, .xls, and more. By converting the attachments into machine-readable text, the team could perform further analysis and processing. To ensure the quality and accuracy of the extracted text, the team employed natural language processing (NLP) methods. Techniques such as *Word2Vec, SpaCy*, and *GreekBERT* were used for text cleaning, normalization, and semantic analysis. Text embeddings were configured to represent the extracted text in a numerical format suitable for machine learning algorithms. These embeddings captured the semantic meaning and context of the text, allowing for effective comparison and similarity analysis.

For each email flagged as important, pairwise similarity scores were calculated between the attachments of crucial emails and other emails using the configured text embeddings. This process identified attachments that exhibited high similarity to the crucial emails. The five most similar attachments were selected for further review. Attachments belonging to unread emails were given priority, as they were likely to contain crucial information not yet analyzed. These attachments were returned and suggested for manual inspection by HCC employees, reducing the number of attachments requiring review and focusing on potentially significant evidence.

2. Graph Analysis of Corporate Entities in Emails

Graph analysis was employed to analyze the corporate entities involved in the email communication. For each email, the To/From/CC/BCC fields were extracted to identify the members involved in the email thread. Entity analysis techniques were applied to accurately identify and classify the entities. The analysis also detected aliases and variations for each entity, enhancing the accuracy of subsequent analysis steps. The entities and emails were represented as nodes and edges, respectively, in a graph structure. Each email thread connected two or more nodes, creating a network representation of the communication patterns among the corporate entities.

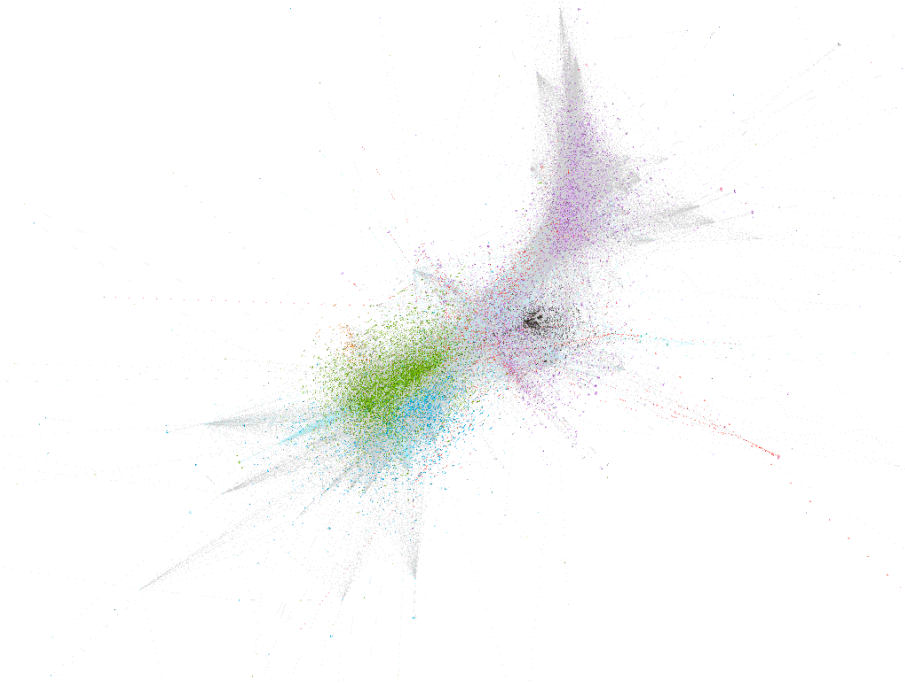


Figure II: The communication graph using the Yifan Hu layout. Each financial institution is assigned a colour so that communities are visible.

Various centrality measures, such as eigen centrality and betweenness centrality, were utilized to identify nodes with high influence on information flow within the graph. Unread emails were also taken into consideration during node selection. The emails associated with the selected nodes were flagged for further review by HCC employees, as they were likely to contain crucial information or contribute to the investigation.

3. Neural Network for Email Body Classification

A neural network model was employed for email body classification, which helped determine the relevance and importance of each email in the investigation. To prepare the email bodies for classification, text cleaning, and normalization techniques were applied. This ensured consistent and reliable input for the neural network model. A transformer ensemble model was employed to characterize the emails as 'Non-Applicable' (0) or 'Crucial' (1). The model leveraged true labels from emails previously read by HCC employees as part of the training data. By applying the ensemble model, emails were classified based on their relevance to the investigation.

By combining the outputs of the neural network model with the results from attachment analysis and graph analysis, a subset of suspicious emails emerged. The neural network's classification helped filter emails deemed crucial (labeled as '1'). Through the collective application of these techniques, the number of emails that had to be manually reviewed by HCC employees was significantly reduced allowing for a more efficient and targeted review process.

III. New workstreams

A. An AI-enabled documental retrieval system with semantic search

To streamline document retrieval and knowledge sharing, we plan to develop a vector database that utilizes a Large Language Model with Legal Expertise (LLM-LE) on top. This system will (i) perform semantic searches to retrieve semantically relevant documents, such as past cases and court decisions and (ii) potentially offer legal insights for an ongoing case based on the retrieved documents. By harnessing the power of NLP and machine learning, the HCC can enhance its research capabilities, access valuable precedents, and strengthen its ability to enforce competition laws.

B. Developing tools for social media and company monitoring

Monitoring social media serves multiple purposes for the Hellenic Competition Commission (HCC). Firstly, it aids in detecting potential antitrust violations by identifying suspicious activities that warrant further investigation. This includes monitoring social media posts, company updates, accounting reports, and investor updates for any indications of anti-competitive behavior. Secondly, social media monitoring helps in gathering evidence for antitrust cases. Additionally, social media monitoring allows the HCC to identify consumer concerns and gain insights into consumer sentiment. By monitoring social media conversations, the HCC can

understand consumer perspectives on various factors such as pricing, product quality, and access to goods and services. This information helps the HCC identify potential issues that may require intervention or further investigation to protect consumer interests and ensure fair competition. Lastly, social media monitoring plays a role in monitoring compliance with antitrust regulations. By monitoring social media posts and communications, the HCC can identify potential violations and initiate investigations accordingly.

C. An AI-revamp of the HCC's existing whistle-blower platforms

Integrating AI-based chatbots into the whistleblower reporting process offers several advantages. By posing specific questions, the AI-based chatbot can effectively guide whistleblowers, ensuring that they provide comprehensive information about alleged anti-competitive behavior. The chatbot's adaptability allows it to tailor its inquiries based on the whistleblower's responses, fostering a more interactive and personalized experience. This adaptive approach helps uncover additional details that may have been overlooked in a simple text submission, enhancing the quality of the information collected.

Moreover, AI-based chatbots possess the ability to seek clarifications and pose follow-up questions as necessary, ensuring the clarity and accuracy of the gathered information. Furthermore, chatbots exhibit multilingual proficiency, enabling them to comprehend and process diverse languages, even within the same sentence. This linguistic versatility facilitates communication with whistleblowers across various language backgrounds, accommodating their preferred means of expression. Additionally, leveraging the information provided by the whistleblower, AI-based chatbots can autonomously categorize and prioritize reports, eliminating the manual burden of tagging and archiving for HCC staff. This automated process streamlines the handling of whistleblower reports, allowing for more efficient and organized management.

D. AI-Assisted Web scraping

Generative AI can play a vital role in assisting with web scraping tasks, enabling the HCC to streamline the data collection and integration steps and add more sources of internet data. Models can be trained to understand the structure and content of web pages, facilitating the extraction of relevant data. By utilizing NLP, AI can accurately parse information from complex web pages, including textual data,

images, tables, and more. Additionally, AI can assist in ensuring the quality and integrity of scraped data.

E. Automatic Inquiry Assignment and Prioritization with NLP

The Directorate-General must, among other duties, assign inquiries to the relevant departments within the HCC. This is a time-consuming and error-prone process, which can be significantly reduced by leveraging AI to assign inquiries to their relevant Directorates and Units automatically. Models can be trained to understand the content and context of inquiries received by the HCC and extract key information. The subject matter, industry sector, and complexity of each inquiry can be determined, and it can be intelligently routed to the appropriate departments based on their expertise and workload. Finally, by considering factors such as caseload, available resources, specialization, historical data and user feedback, the HCC can continuously improve the model's performance in inquiry assignment.

IV. Conclusions

Integrating data science and data scientists in competition law enforcement and the day-to-day work of competition authorities presents one of the important challenges for competition law enforcers in the years to come. Advanced data analytics and generative AI may revolutionize the work of competition agencies and may enhance their capabilities to detect and punish anticompetitive activity. At the same time, these new tools offer significant capabilities in terms of market mapping and market monitoring, thus enhancing the ability of the competition authority to use a variety of tools (infringement decisions, settlements, sector inquiries, market investigation references, soft law approaches) and to make a deliberative and flexible (responsive) choice from the various regulatory strategies at their disposal. The Report presents the various workstreams of the HCC in-house Data Science team and explains how data science tools were employed in several competition law investigations.

Kenya
Competition Authority of Kenya

Businesses continue to leverage information and communication technology tools globally, especially through digitization, in order to be more efficient and reliable in their processes.[29] This brief demonstrates how integration of customer relationship management ('CRM') and enterprise resource planning ('ERP') systems have improved operational efficiency in competition law enforcement in Kenya. These systems have become key in storage, aggregation and referencing of agency-wide records and information, and as a result reduced filing costs and turnaround time by up to 50 per cent. The installation and management of the systems has enabled uninterrupted service delivery in the face of adversity and shocks associated with the global COVID-19 outbreaks. This brief provides a testimony of how competition agencies can take advantage of investment in and installation of technology-enhanced effectiveness and improve the operational efficiency of their regulatory efforts. A key insight is that upskilling human resources is critical for the implementation of computational tools. Furthermore, the adoption of such tools is also impacted by the recent development of data protection and privacy laws in the digital economy landscape. The data and information aggregated through digitization allow for evidentiary, coherent and predictable competition law enforcement.

I. The automation of complaints/ case processing

In recent years, businesses globally have leveraged information and communication technology tools in order to be more efficient and reliable in their business processes. The COVID-19 pandemic heightened the need to enhance the use of digital tools, including in the context of the detection, prevention and deterrence of anticompetitive conduct by competition agencies.

The Government of Kenya clearly notes the potential of digitization to boost the growth of African economies. The need for a regulatory framework that facilitates digitization stems from the fact that digitization boosts investments, enables product innovation and the development of data-based services.[30] In this regard, the

[29] This note was authored by Adano Roba, Gideon Mokaya & Mugambi Mutegi. Correspondence: aroba@cak.go.ke

[30] The Republic of Kenya, *Digital Economy Blueprint: Powering Kenya's Transformation*,

Competition Authority of Kenya ('the Authority') cognizant of the need to enhance efficiency of its operation, for both internal and external stakeholders, operationalized the Case Management System ('CMS') in 2020, after almost four years from conceptualization. The CMS has automated the processing of matters related to merger filings, restrictive trade practices and abuse of buyer power cases, as well as consumer protection complaints.

The agency's digitization process commenced in 2018 when the Authority converted all its paper files into digital records and archived them in relevant streams in the CMS and integrated them with the Document Management System ('DMS') for easy referencing. The digitalization process has also enabled the case handlers to access the system remotely, and has rendered compliance with the data protection laws easily attainable.[31]

The introduction of electronic DMS records was a significant milestone since any undertaking's submission received in hardcopy is now scanned and uploaded by the Authority's registry personnel to be assigned to a case handler for processing. The CMS and DMS systems completed the cycle by facilitating undertakings/stakeholders to upload digital submissions remotely, from any geographical location both nationally and globally, through the Authority's Online filing ('E-Filing') Portal.[32] Filings made through this online platform are automatically assigned to CMS files and the registry officials are notified. Further, any attachments, including evidentiary information if it pertains to a complaint or merger application form from external parties, are automatically captured in the system.

This process also ensured the full digitization of the entire Authority's operations with the support of functional units (departments) through the enterprise resource planning ('ERP') system. This system caters for, among others, human resource productivity monitoring, procurement, electronic submission and review of job applications, application of various types of leave, and non-technical processes that are critical for the smooth functioning of the Authority.

---

https://www.ict.go.ke/wp-content/uploads/2019/05/Kenya-Digital-Economy-2019.pdf (last accessed June 11, 2023).

[31] The National Council of Law Report, *The Data Protection Act No. 24 of 2019*, https://www.odpc.go.ke/download/kenya-gazette-data-protection-act-2019/?wpdmdl=3235&refresh=64834673538de1686324851 (last accessed June 11, 2023).

[32] Competition Authority of Kenya, *E-filing Portal* https://competition.cak.go.ke:444/ (last accessed June 11, 2023).

The CMS and the DMS systems are comprised of (dashboard and workflows) that have the following main features:

- The DMS contains the record of all saved files and their exact location with an instantaneous search engine;
- There is a file management module integrated with an administration and information upload interface used by the Registry Team;
- CMS interfaces are used for instructing case handlers (legal and economic analysts) to determine the progress of each file assigned to them;
- The local control interface enables the Authority's senior Management to be notified of all files assigned to their domain, their status, and case officers interacting with them;
- The general control interface enables staff to know all the files assigned to each department of the Authority, the progress, and case officers who interact with them;
- Control users have an overview that includes the following information and functions:
    - Real-time reports which include the cases handed down in specific number of days and communicated to relevant staff;
    - Flagging case files whose processing dates are maturing and due to expiration or require urgent action;
    - An overview of files initiated at the Registry, officers handling them at the time and their review status;
    - The number of cases in process, classified into those with a final decision or resolved, and those that are still being processed;
    - The timeline of each system file, with details of the stages it has gone through during the processing and the respective dates and duration of each stage.

The CMS and the ERP systems have enabled faster operations and seamless contact with the stakeholders and customers. Filing for a Merger, for instance, before the implementation of the current systems, involved the applicant printing documents, binding and hand-delivering them via ordinary post to the Authority, within the official working hours. This was not only costly, but also time-consuming. Following the automation of the Authority's services, stakeholder feedback indicates that filing costs and turnaround time have been reduced by up to 50%. Applications can be made round the clock since September 2019. Today, all that a party or stakeholder needs is access to internet from any part of the world to file a merger.

In promoting transparency in public procurement, the Authority has embraced e-procurement, whereby bidders submit their bids online and can track their progress. The Authority's e-procurement portal has further enriched customer experience by providing an interactive system where surveys are conducted in a rather cost-effective manner that would otherwise have been costly and involved inconveniences.

In a nutshell, the Authority has saved 90 per cent of operational costs due to digitization, including printing and human capital. The turnaround time was reduced from five to two days, increasing efficiency and speed. Internally, processes such as financial and human resource management have also been automated, resulting in efficiencies such as shorter turnaround time in attending to matters concerning both external and internal stakeholders. In summary, the deployment of the CMS and DMS systems have enabled the Authority execute its mandate in a more cost-efficient manner and facilitated a faster pace for the resolution of cases, thereby enhancing customer satisfaction and raising the profile of the Authority as a well-functioning Government agency.

II. The use of digital tools to detect, investigate and deter anticompetitive practices

The proactive digital tools used by the Authority consist of installations that are relied on in cartel investigations, in accordance with the provisions of the Competition Act No. 12 of 2010.[33] Furthermore, the enforcement of public procurement laws is also digitized in order to ensure that the goods/services procured are adequately priced and of good quality.

In order to enhance the quality and potency of its cartel investigations, the Authority has competitively contracted external digital consultants with specific hardware and software skills and knowledge. Digital tools are relied on in the investigative process to collect, process, and analyze digital evidence, therefore improve data gathering during dawn raids and the subsequent data analysis. The use of such modern tools facilitates the efficient analysis of data running into tens of terabytes. The software further facilitates the generation of reports that enable better interpretation of the evidence resulting from the data collected.

[33] The Republic of Kenya, *The Competition Act of Kenya No. 12 of 2010*, https://www.cak.go.ke/sites/default/files/Competition_Act_No._2012_of_2010.pdf (last accessed June 11, 2023).

The hardware used by forensic experts, in liaison with the Authority's staff, include TX1 Forensic Imagers which greatly reduce the time taken to develop forensic images of retrieved data. Further, the Detego Forensic Imager and Analyzer permits access to and reading of the data in storage devices without compromising the integrity of the data. The analyzer has the capability to read from various storage media (hard drives, USB and servers) that contain the digital evidence. It includes writing blockers to guarantee the protection and integrity of the files subject to analysis, thereby preserving the chain of custody of evidence. The Authority also employs an FTK Analysis Toolkit with the capacity of forensic imaging of data in storage devices, either as one file or in segments that may be reconstructed later for analysis.

Additional equipment deployed by the Authority's external consultants in recent dawn raids (cartel investigations) includes the Fred's Mobile Phone Data Extractor. The tool, which the Authority made use of for the first time in December 2021, has proven useful for assisting in the search for keywords that point to concerted practices among competitors, thereby helping build a strong case against the undertakings under investigation. The extractor also facilitated the acquisition of relevant information from the mobile phones of key suspects. The Authority is in the process of procuring these tools in order to build an in-house forensics laboratory and expertise, in order to cut cost and at the same time build in-house competency of the internal users. These tools are expected to improve the Authority's ability to obtain and analyze data pertaining to the investigation of anti-competitive practices. Further, case-handlers have indeed received training from various forensic experts and are therefore able to competently deploy the tools to achieve the desired objectives of data retrieval and analysis.

In collaboration with the public procurement regulator—Public Procurement Regulatory Authority ('PPRA')—the Authority is in the process of developing a toolkit for the detection of bid rigging through the analysis of key information, and this entails uploading tender documentation.[34] Using algorithms, the investigation tool will detect red flags in key areas of potential concern such as the number of bidders, pricing patterns, document origin and staggered bids. The existence of a red flag, or a combination of several indicators, could signal bid-rigging, thereby warranting a closer review of the applications, including physical assessment. Bids

[34] The Republic of Kenya, *The Data Protection Act No. 24 of 2019*, https://www.odpc.go.ke/download/kenya-gazette-data-protection-act-2019/?wpdmdl=3235&refresh=64834673538de1686324851 (last accessed June 11, 2023).

that score highly in this scale are immediately flagged for more advanced investigations.

The collaborative partnership with the public procurement regulator is borne of the fact that the public procurement of goods and services accounts for approximately 40 per cent of the Government of Kenya's budget annually. Therefore, efficiency in this sector is key to converting fiscal plans into results. Once successfully completed, this project will enable the Authority to detect and resolve allegations of bid rigging with relative ease, thus increasing the scope of surveillance of public procurement. The net effect of bid rigging is that purchasers--national and regional Governments--end up paying much more than the competitive price for services whose quality is often not guaranteed. Collusive tendering usually also stifles innovation.

III. Use of Digital Tools to Increase Agency Effectiveness

In relation to the analysis of market inquiry data, like the concluded Digital Credit Market Inquiry, the Authority used an open source data analysis tool--R Analytics. This software was cost effective and provided an opportunity for analysts to master various commands and stay up to date with updates and new modules with expanded versatility. However, the downside of this software is that it has limits on the volumes of data that can be processed. This is unlike software such as the statistical program for social sciences ('SPSS') which can process considerably larger volumes.

The Authority has also incorporated open source Knowledge Management tools to manage both the resource center and the digital content. These are KOHA and D-Space. KOHA, which is utilized by over 3,000 libraries across the world, has given the Authority access to a wide array of relevant and up-to-date resources and information from the library. On the other hand, D-Space preserves and supports easy and open access to digital content including text, images, moving images, .mpegs and data sets available within the Authority. All Authority staff have access to the data repository that supports the development of various reports and advisories.

IV. Conclusion

The relevance of technology for competition enforcement has become evident for the Authority, therefore the investment in the development and implementation of a case management system, acquisition of digital forensic tools and usage of off-

the-shelf software are paying off. It is not only about being able to understand how disruptive technologies work but also to fulfill the legal mandate of the Authority effectively and efficiently by drawing on resources and tools that support priority activities and enhance the timely resolution of cases.

The use of computational tools like algorithms and artificial intelligence are still in their nascent stages of use by competition agencies in developing countries. However, the Authority takes a proactive approach in upskilling its case handlers regarding the latest technology, knowledge and practical skills to ensure the officers are able to keep pace with the dynamism and rapidly changing the face of competition law enforcement. Mindful of this, the Authority aspires to learn from the expertise of other agencies through the Stanford Computational Antitrust Project.

Luxembourg
Luxembourg Competition Authority

The present contribution is an update to the one provided last year. In this logic, the context and activities described in sections I and II remain unchanged and are complemented by a brief progress report in section III.

I. Context

The Luxembourg Competition Authority ('Authority') is the competition authority of Luxembourg. Until now, its missions have been the enforcement of national and European competition law. However, its tasks have grown. For example, the Authority oversees the enforcement of the directive 2019/633 'Unfair Trading Practices (UTP)' and regulation 2019/1150 'Platform-to-Business (P2B).'

In order to prepare itself for these new missions and to react to the ongoing digitalization, the Authority set up a new unit dedicated to platform markets and data analytics. In a first stage, two people will be working in this unit and the immediate tasks in the short term will be:

- To monitor the local platform economy and handle complaints in the framework of the 'Platform-to-Business' regulation;
- Assist in the analysis of competition cases in the digital economy;
- Contribute to more general activities of the Authority, like sector inquiries through improved data analysis.

II. Activities of the new unit

A. Increase current know-how

In a first phase, the main goal of the new unit will be to increase its capabilities. Before hiring additional staff, the current employees will increase their expertise through trainings in fields of statistics, data science, and platform economics. A second way to increase the expertise of the Authority will be the collaboration with other authorities, especially in the framework of the ECN network. Finally, the unit will reach out to the academic field to explore possibilities of collaboration. In this regard, the participation of the Authority on the Computational Antitrust Project is a first important step. Additionally, the Authority will reach out to other institutions

like the University of Luxembourg to exchange on ongoing legislative evolutions (*e.g.,* DSA and DMA) and to learn more about computational tools.

B. Put in place specialized tools

In a second phase, the aim will be to put in place specialized tools to improve the Authority's functioning. Possible projects in this regard would be the following:

- The use of sophisticated 'eDiscovery' software: This will improve the handling and analysis of firms' documents. The number of documents a competition authority has to analyze in a given case is growing rapidly. An effective analysis of these documents can no longer be done manually. Therefore, it is crucial for the Authority to improve in this area through sophisticated eDiscovery tools, which allow a rapid identification of those documents that are important in an ongoing investigation. Such tools will rely on AI techniques such as Natural Language Processing (NLP). Given the complexity of such tools, they will likely be acquired rather than developed in house.
- Monitoring of Terms and Conditions of platforms: In the framework of the 'Platform-to-Business' regulation, platforms' terms and conditions have to respect certain obligations. To be as efficient as possible, an automated tool will be developed to go through the terms and conditions of firms and identify potentially problematic cases. In this way, the Authority will be able to ensure a pro-active market monitoring rather than reacting solely on complaints.
- Market monitoring: Similarly, automated tools could also be put in place to monitor markets in general. Be it the analysis of prices on specific markets to detect potentially anti-competitive behavior or the automated analysis of press articles.

As experience will grow over time, the unit will readapt its goals and hire additional people according to the identified needs. In all likelihood, these new profiles will stem from the fields of econometrics, data science or IT.

III. Progress report for 2022-2023

The activities set out last year remain the Authority's plan to set up a unit focused on the analysis of platform markets. Currently, the unit's activities in this regard are mainly to increase its current expertise (cf. section II.A.). The projects set out in section II.B. could so far not be put in place. A higher-than-expected need for the unit's staff to contribute to antitrust investigations was the main reason.

Mexico
Federal Economic Competition Commission

The Mexican Federal Economic Competition Commission ('COFECE') has actively engaged in efforts to approach the challenges posed by digital markets and achieve a high degree of digitalization within our agency, particularly since we published our Digital Strategy back in 2020.[35]

We define digitalization as a process that seeks to introduce, expand, and standardize the use of computational tools and technology-driven initiatives within different activities to enhance our institutional capabilities. Thus, the process of digitalization within our agency has an impact in processes, activities, as well as in detecting and enforcing anticompetitive conducts.

As part of our engagement with digital markets and digitalization we would like to refer to a couple of international cooperation initiatives that we believe will provide strategic insights and platforms to exchange best practices on these matters.

First, COFECE has been working with fellow competition agencies in the Americas to launch the Group of Competition Agencies of America (GrACA).[36] Within GrACA, COFECE and the Brazilian Administrative Council for Economic Defense (CADE) co-chair an agency-only working group on digital markets and digitalization. This working group also includes two sub-groups, one on digital markets and the other on digitalization within the agencies. The purpose of these is to exchange best practices, ongoing projects, and identify potential agency cooperation initiatives.[37]

Second, in November 2022, a series of working meetings related to technology and digitalization were held between members of the United States Federal Trade Commission (FTC), the Department of Justice Antitrust Division (DoJ), the Canadian Competition Bureau (CCB) and COFECE. This allowed our agencies to share

[35] Available in Spanish at COFECE, *Estrategia Digital*, https://www.cofece.mx/estrategia-digital-cofece/ (last accessed June 7, 2023).

[36] The acronym stands for Grupo de Agencias de Competencia de América in Spanish, *i.e.* GrACA. This is an initiative that relaunches an interagency cooperation forum formerly known as *Regional Competition Center*.

[37] The General Directorate of Digital Markets leads our efforts in the digital markets working subgroup and the General Directorate of Market Intelligence coordinates COFECE's contributions in the digitalization within agencies working subgroup.

highlights from digitalization processes and tech- focused projects, identifying potential areas for future cooperation. Therefore, COFECE's engagement with digitalization, which includes the use of computational tools, has featured an active agenda of international cooperation.

In the following section we will briefly discuss specific examples of computational tools that have been used in both internal processes and activities. Furthermore, we will refer to some of the applications of computational tools to detect and enforce anticompetitive conducts.

I. Use of computational tools in internal processes and activities

In January 2020, COFECE launched SINEC, which is a system that allows us to conduct all merger review processes electronically.[38] Some of the advantages that result from this system is the possibility to notify an intended merger remotely, having continuous and uninterrupted access to the case files, and expedite the overall merger analysis.[39]

In addition to the latter, COFECE has also digitalized and automatized additional processes and services which include case management tools such as the electronic reception of documents related to ongoing cases, including the submission of formal complaints of potential anticompetitive conducts. We also are currently able to conduct remote compulsory interviews, among other investigative tools.

Currently, we are working on strategic projects that provide our digitalization agenda with even greater relevance. For example, we are in process of introducing the use of advance cognitive services including GPT-4 and Generative Artificial Intelligence (GAI) in different internal processes and services. This is possible because we have managed to succeed in several initiatives that have included the adoption of cloud-based infrastructure arrangements, intensive digitization of information, development of semi-automatization and automatization processes, among other.

[38] Sistema de notificaciones electrónicas (in Spanish). For more information, *see* COFECE, *Sistema De Notificaciones Electrónicas*, https://www.cofece.mx/wp-content/uploads/2020/03/InfografiaSINEC.pdf (last accessed June 7, 2023).

[39] COFECE's Board issued new guidelines and regulatory provisions that legally enabled the use of electronic and digital means in several of our formal processes.

It is important to note that as part of our 2023 yearly projects, COFECE is working on developing processes and tools to further enhance the optimal use of different information sources collected as part of ongoing antitrust investigations. We consider that this approach would contribute to improve the effectiveness and efficiency of our processes and activities.

II. Use of computational tools in detecting and enforcing potential breaches of Mexican competition law

COFECE's Investigate Authority (IA) has the General Directorate of Market Intelligence (GDMI or Intelligence Unit) as part of its structure. The GDMI detects potential anticompetitive conducts in the Mexican market to develop ex officio investigations. For this purpose, the Intelligence Unit uses a wide variety of tools that range from Open-Source Intelligence (OSINT), Human Intelligence (HUMINT), data and economic analysis, to mention a few.

Since its creation in 2014, the Intelligence Unit has built capacities both in terms of human and material resources to be able to improve the effectiveness of its detection activities. Moreover, the GDMI also provides specialized services to other areas of the IA that investigate cartels, abuse of dominance and barriers to competition.[40]

In this regard, the GDMI has developed several projects that rely on computational tools for the detection and enforcement of anticompetitive conducts. For such purpose, we could divide these projects in two broad categories: i) Data collection-processing tools, and ii) Data analysis tools.

In the first category, we could include tools that have been developed for automatized data collection and processing tasks, which feature a variety of algorithms capable of web scrapping, data extraction from different formats, data base integration from both structured and unstructured data origins, to mention a few. These tools have been built to enhance the efficiency in collecting large volumes of information, particularly quantitative, but are also useful for OSINT tasks.

[40] The Investigative Authority is formed by five General Directorates: Cartels (Dirección General de Investigación de Prácticas Monopólicas Absolutas), Unilateral Conducts (Dirección General de Investigaciones de Mercado), Regulated Markets (Dirección General de Mercados Regulados), Coordination Office (Oficina de Coordinación), and the Intelligence Unit.

In the second category, the GDMI has developed a series of specific-purpose tools for both markets and cases. For example, we have developed a public procurement screening tool that is useful to detect potential anticompetitive patterns (both price and non-price related) in Mexican federal government procurement data. Our screening tool relies on the use of computational tools and provides a dashboard that facilitates analysis tasks. From this tool we have managed to build several major public procurement case leads.

Moreover, we have also developed tools useful for the detection of potential unlawful mergers, for the analysis of alleged unilateral conducts and barriers to competition. For example, the GDMI built a tool for geographic relevant market definition that has been applied in gasoline and LP gas cases with the purpose to define potential relevant markets using both pivot and non-pivot methods.

These couple of examples provide a general overview of the extent in which we have adopted computational tools as part of detection and enforcement activities within COFECE. We believe that these would also be included as part of our digitalization efforts and intend to enhance our projects with new applications such as GAI. On this matter, we are also exploring using GAI in the analysis of digital evidence acquired using IT forensic tools during dawn raids with early promising results.

At COFECE we are committed to continue our efforts to achieve a profound digitalization within our agency, which will be inherently related to the application of computational tools for both internal activities as well as enforcement tasks.

Poland
Office of Competition and Consumer Protection

In recent years, AI has revolutionized the way organizations operate, providing solutions to problems that were once deemed very complicated or even unsolvable. The Office of Competition and Consumer Protection ('UOKiK') has implemented several AI tools to detect anticompetitive practices, enforce the law, and improve work processes. These tools have proven to be incredibly beneficial, helping the Office to operate more efficiently, effectively, and with greater accuracy. UOKiK has been able to increase the quality of its work overall and streamline its operations with the use of AI. Below is a description of most relevant AI tools implemented by the Office and their impact on the organization.

I. Bid Rigging Detection

In the area of competition protection, the UOKiK is developing a method for the automated detection of collusive bidding with the use of statistical tools. In recent years, we have conducted activities that result in analyses of data on large infrastructure projects (road and railway projects). One set contains data provided by the General Directorate for National Roads and Motorways. The database contains information on 3174 offers submitted. The second set contains data provided by the Polish State Railways. The database contains information on 1986 submitted offers.

On the basis of this data, we have calculated measures allowing to characterize the structure of the market and its dynamics, *i.e.,* measures of concentration, HHI, in general and regional terms. The analysis of the market structure does not show any alarming trend, the market is not dominated by one company or by a group of companies. Market concentration neither increases nor decreases appreciably. Its volatility is considerable, which is common in bidding markets where winning one large tender can lead to large changes in market share.

The first measure that was analyzed was the convergence of the price with the estimated value--it is considered that the price should be lower than the estimated value determined before the announcement of the tender by its organizer. If it is not, it may be the result of inflating the price by non-competitive behavior of bidders. The price-to-estimate ratio, measured as the ratio of the winning price to the estimate, was calculated. This measure exceeds 1 if the most advantageous price exceeds the value of the order estimated by the organizer.

Next, the volatility index was calculated. This indicator is determined as a relative standard deviation, *i.e.,* the standard deviation of prices divided by the average of the prices of bids submitted in a given tender (only tenders with more than 2 bids were taken into account). Then, procedures characterized by particularly high or low price volatility were identified.

The asymmetry of price distribution in individual tenders was also analyzed using the skewness index. Positive skewness means that the majority of offers are above average. This may indicate artificially expensive losing bids--the bids that are supposed to lose are set at high prices to make the winning bid seem more attractive against them. Another possible explanation is that the winning bid was underestimated, *e.g.,* as a result of a mistake or to gain needed experience.

Suspicions may also be raised by a particularly large gap between the first (best) and the second-best offer. To take into account the specificity of the market, the range index was calculated by dividing the range between the first and the second offer by the standard deviation of the offers excluding the first. This measure is intended to detect a situation in which the best price differs significantly from the others, whose values are rather concentrated.

It was also examined whether reciprocity can be observed between the companies whose bids were withdrawn and those which won the tender as a result (A withdraws and as a result B wins, then at some later date the situation is reversed - B withdraws, allowing A to win with a higher price).

In addition to situations of direct reciprocity between winners and withdrawers, more complex arrangements should also be considered. When firms strategically withdraw bids in multiple tenders, a deal is possible such that firm A withdraws in tender one, firm B wins that tender, withdraws in tender two, which is won by firm C, which in turn withdraws in tender three, so that company A wins. This situation can be called a cycle. Cycles can also be longer than three elements.

However, it must be taken into account that if these companies participate in tenders not on their own, but in consortia that are not permanent (in individual tenders, a given company participates in consortia of various composition), then the implementation of such an agreement would be much more difficult, and therefore there are doubts as to whether this is intentional.

II. Preliminary assessment of the risk of tender collusion

Regardless of the works described above, UOKIK conducts research and development work on the implementation of mechanisms for a preliminary assessment of the risk of tender collusion. Although we have a prototype tool that allows initial data analysis, the basic problem is the availability of high-quality data on the public procurement market. In order to develop appropriate databases, UOKiK cooperates with the Public Procurement Office, working on the use of data collected under the eProcurement (eZamówienia) system.

III. ARBUZ—AI-powered assistant detecting abusive contract clauses

Employees of UOKiK, react on received complaints or take action on the basis of market research, media reports and information from other sources. Many investigations are precedent-setting, requiring a multidimensional analysis of a particular problem. However, some activities are rather repetitive and laborious at the same time. Such tasks include the analysis of standard contracts in order to identify abusive clauses. UOKiK is obliged to ensure that contracts used in mass trade do not contain provisions detrimental to the interests of non-professional users. The search for such provisions consists in reading multipage standard contracts and then assessing which parts of them may infringe consumer rights. It is a time-consuming process. For this reason, UOKiK carried out an EU-funded project 'ARBUZ,' aimed at automating the process of analyzing standard contracts.

Employees involved in detecting illegal clauses can log into ARBUZ and start searching for model contracts. Already at this stage, it uses AI to assess whether the document found falls into this category. The program can also be fed with contracts received along with consumer complaints or obtained from other sources, for example, submitted by the investigated company at the request of UOKiK. ARBUZ is equipped with an OCR system, which allows you to work not only on editable documents, but also on pdfs, graphic files and scans in various formats.

ARBUZ reviews the contract selected by an employee. Using the slider, you can set the sensitivity with which the analysis is to be performed—*e.g.,* 5, 10 or 20 illegal provisions are to be found. The system compares the contract uploaded manually or found with the help of the crawler, with clauses in the database. In two windows in the interface you can see the provisions which according to the robot may be unfair. In the first window the user sees the contract with automatically found clauses marked in color, and in the other, the same clauses with information on how likely they are, according to the system, to be unfair. ARBUZ does not just make a simple comparison

of the provisions of contracts with database entries but uses intelligent algorithms to recognize the meaning of complex sentences written in legal language. It works at the level of an intelligent assistant--it is more than a tool but at the same time it is not allowed to make independent decisions.

At the end of the process, however, there is a human case handler who decides how to deal with the results of the ARBUZ's work. First of all, he or she may recommend bringing charges against the entrepreneur or sending a letter requesting a change in the standard contract.

IV. Dark Patterns

One of the current online threats, against which practical standards of combating have not yet been developed, are dark patterns--user interface design choices that benefit an online service by coercing, steering, or deceiving consumers into making unintended and potentially harmful decisions.

UOKiK has initiated an EU-funded project to develop a methodology for conducting proceedings on dark patterns and explore AI deployment opportunities for consumer protection authorities through the delivery of a Proof of Concept of an AI-powered tool for detecting dark patterns as well as the research on the latest AI and machine learning trends. Websites of online stores and other online service providers will be screened and specific online market sectors with the highest dark patterns occurrence will be selected for further analysis.

Next, explanatory proceeding will be initiated on the basis of the findings of an open-source intelligence, which will also include neuromarketing tests will be carried out to examine neurobiological human reactions while being exposed to dark patterns. In the course of administrative proceedings, UOKiK plans to organize an IT analysis of websites, delivered by external experts, and conduct a consumer survey about online purchase experience and the influence of certain dark patterns.

In the next stage, the results of IT website analysis, consumer survey, and neuromarketing tests will be used as input data in the Proof-of-Concept process of a dark patterns AI-detector. The project team will also prepare the main assumptions of the AI tool and an open idea competition will be organized. The contest formula stimulates innovation and allows for the selection of the best contractor on the basis of their real competences rather than experience or market position.

The results of the project will be presented in the form of guidelines for conducting proceedings on dark patterns and a white paper on practical, legal, and ethical aspects of using AI by a public body enforcing consumer law.

Romania
Competition Council

The Romanian Competition Council ('RCC') must manage increasingly large volumes of data that cannot be stored and queried exclusively using traditional RDBMS (relational database management systems). On the other hand, in recent years, there was a clear trend to integrate more and more intelligent processes into applications, while functionality remains at the basic requirement level for ease of use. Also, since the abundance of sources implies a lot of disparate information, apparently unrelated to each other, predictive analyses implemented through Big Data technologies allow RCC to identify connections and patterns of interest, the identification of which would be impossible in the absence of these tools.

The development of the main component of our digital investigative system, the Big Data Platform, finalized at the end of 2021. This component allows the specialized staff of the RCC to access/use data for the initiation and development of cases and market studies based on 5 analysis modules: cartel screening, bid rigging detection, sector inquiries, mergers and structural and commercial links between enterprises.

An important aspect of our digital investigative processes is access to correct and complete data, their processing and interpretation. For these reasons, data acquisition is an important process and allows the downstream components of data processing to perform their functions correctly.

In the data acquisition process, the RCC platform administrators focus on several important aspects:

- picking up frequent data changes to ensure that the data destination is in sync with the content source;
- storing the data as it is retrieved; any data operation (normalization, cleaning, enhancement) is performed later in the loading/storage process;
- obtaining all the necessary metadata from the content source;
- allowing for the possibility to retrieve files stored in the content source.

The main data processes in the RCC BD platform are presented in the following diagram.

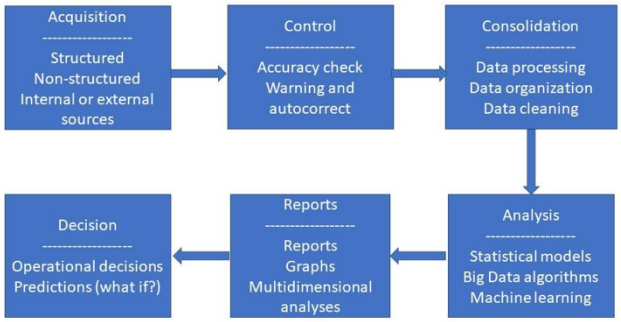


Figure I: data processes

I. Acquisition

The acquisition process consists of data imports from specific sources of internal or external data, representing transactional, operational data, content, references or system-generated data. Data is stored in relational databases, distributed file systems and NoSQL databases.

The external data which is part of our online architecture comes from the National Trade Registry, the tax authority, the electronic public procurement platform, the National Council for Solving Complaints in public procurement and the Ministry of Justice.

Apart from the external sources, the platform also utilizes most of the data RCC already stores since it automated internal workflows and digitized physical archives. The daily press report that RCC receives as part of an external contract is also uploaded to the platform. In addition, RCC already had digital databases which harbor important data, such as the Price Monitor app for fuels and food products and the State Aid Registry.

In the future, if deemed appropriate, an avenue that could be explored is to upload case documents into private and secure workspaces of the BD platform, where more advanced operations could be performed, such as text analytics.

II. Control and cleaning

A large portion of the structured data is imported from external sources. Until now, the digitalization process of Romanian public administration was hindered by a low level of interoperability of public administration IT systems. The existing systems are generally fragmented, being designed and developed in isolation by different state institutions, outside a coherent national framework. Consequently, the external databases providing input to our platform were built without having in hindsight the need to standardize variables. The structured datasets can contain errors, especially if the original data was input by humans and the content was not checked by automatic mechanisms.

Defining an appropriate data structure that connects company ownership and financial data, public procurement datasets and qualitative or processed unstructured data – and makes filtering across these possible – requires a very advanced understanding of each processed dataset. Therefore, during the implementation of the platform, the project team had to analyze the datasets and identify possibilities for designing an optimal structure for our platform, as well as potential errors that might arise when trying to match datasets.

For example, one company can appear under several differently spelled names: short, long, misspelled, etc. Since one of the main functionalities of the BD platform is centering all collected information around companies and their relevant markets, the possibility to match sets of data related to a specific company is an essential part of the database-building process. The companies registered in Romania have a unique numerical identifier allocated by the National Trade Register, therefore a deterministic matching was made using this unique ID to overcome issues related to companies appearing under different names in different datasets.

For certain aspects, code was written to automate the cleaning process. For example, the unique numerical identifiers needed to be standardized since some included the country code (RO) and some did not. Also, the line feed was eliminated from CSV files in which the character fields contained the LF (line feed) character in order to correctly navigate the structure of the received file.

Another matching issue with our imported datasets is the lack of unique identifiers for natural persons who are shareholders or administrators of a company.

Even if the original databases, belonging to the institutions from which we acquire data based on collaboration protocols, include unique identifiers for persons (Personal Numerical Code), compliance with the GDPR policy has allowed us access only to the name, date of birth, city and county of birth in free text format.

The automatic data acquisition processes do not contain matching algorithms for natural persons. Therefore, during specific case investigations, if there are multiple results in a query, RCC experts have to establish if several records belong to the same person and define a 'same person' relationship between the respective records, which may be time-consuming and defeats the purpose of the platform.

The GDPR issue came up earlier as well, at the setup of the platform, when acquiring entire structured databases from other public authorities raised GDPR concerns. RCC was already using an interoperability platform that allowed it to interrogate databases of other institutions. The inquiries were however specific to a particular company, it did not involve large data sets. The Big Data platform, on the other hand, needs access to entire sets of data to screen and issue red flags. To alleviate GDPR concerns as to why access to privileged information of so many companies is needed in one go, the RCC BD team had to carefully justify the process and explain the legal underpinnings. RCC has currently made steps toward amending national legislation in order to overcome these hurdles.

III. Conclusion

Maintaining and improving data sets is an ongoing exercise for which long-term commitment is needed and resources need to be allocated. Nevertheless, one of the biggest and immediate added values of the RCC BD platform is the quick access to integrated administrative datasets.

Singapore
Competition and Consumer Commission

I. Document Generation Tool (DocGen)

As part of the merger review process, case teams have to redact confidential information claimed by different parties in a document. In scenarios where they have to share the redacted document to various parties, they will have to prepare different versions of the redacted document such that each party can only view the confidential content claimed by them but not other parties. Thus, if there are many parties involved, case teams will have to go through the tedious process of 'tagging the confidential contents of different parties,' 'redacting' and 'preparing' different version of the redacted documents. This manual process is time-consuming and prone to human errors where confidential information claimed by Party A is not redacted in the version of the redacted document shown to Party B.

To address this problem, the Competition and Consumer Commission of Singapore ('CCCS') developed a document generation tool (DocGen) to reduce the time, effort and possible human error when preparing different versions of redacted documents for different parties. The case teams will use DocGen to 'tag' confidential information claimed by different parties in the document, and after completing tagging all the confidential information in the document, the case team will then use DocGen to generate redacted document for each party (or parties) easily.

With DocGen, the time and efforts to prepare different versions of the redacted document is reduced significantly and human errors are minimized as it allows CCCS to have a more structured and standardized way of redacting documents.

II. Robotic Process Automation using UiPath

CCCS also embarked on its Robotic Process Automation ('RPA') journey using UiPath to develop robots to automate certain manual processes.

The first robot developed is the Complaint Data Entry robot. Prior to the introduction of the robot, CCCS officers had to sieve through complaint emails daily and manually create each complaint record in the CCCS complaint management system. This manual process is time-consuming and prone to data entry error. To

overcome this issue, CCCS developed the Complaint Data Entry robot to automate the repetitive process. This automation helped to significantly reduce the time for data entry as well as human errors.

Separately, CCCS also developed an in-house Excel macro called 'Workload Management' to collate data from the various projects or cases undertaken by case officers. This is a dashboard tool for CCCS management to have an oversight of the workload of case officers. CCCS also developed a robot to automate the data refresh and extraction and the generation of dashboard reports without human intervention.

CCCS recognizes the potential benefits of RPA and will continue to harness the technology to automate corporate processes, especially those which are repetitive, tedious, and manual in the area of human resources, procurement, and finance.

Slovakia
Antimonopoly Office

The aim of this article is to describe the unsuccessful attempt of the Antimonopoly Office of the Slovak Republic ('the Office,' 'the Authority') to apply a specific machine-reading tool in order to collect data from published decisions of the Office and the lesson learnt from this attempt. In light of recent advances in digitization, web-scraping and web-crawling technologies, the Authority has considered the implementation of advanced tools for machine reading of digital content admissible on the internet in order to analyze and detect suspicious anticompetitive behavior. The Slovak language, as well as the legislation, are specific to Slovak business environment. The Office therefore launched a Proof-of-Concept project to measure the accuracy of such tools. Following the initial analytical phase, the Authority launched a preliminary market consultation to formalize the assignment and procure a contractor for the aforementioned Proof-of-Concept project. This preliminary market consultation resulted in an indicative quotation, on the basis of which the Authority concluded that for smaller authorities such tools are relatively expensive compared to manually carrying out certain 'data scraping' processes.

The Office is the independent central state administration body of the Slovak Republic for the protection of competition and state aid coordination. The Office intervenes in the cases of cartels, the abuse of a dominant position, vertical agreements, it controls mergers, which meet notification criteria moreover, assesses the conduct of state and local administration authorities if they restrict competition by their conduct. It ensures the protection of competition in the area of state aid as well.

All decisions of the Authority have been published on the Authority's website. Due to historical developments and changes in the scope and format of metadata associated with the decisions, the Authority faced the need for massive metadata enrichment and cleansing. The enriched database of the Authority's decisions could be the source of valuable data for various analyses of these decisions. In terms of content, the Authority could thoroughly analyze the fines or other sanctions imposed (*e.g.*, prohibition to participate in public procurement), identify the parties to the administrative proceedings; find out whether specific legal institutes (such as leniency program, settlement) were applied or the anti-competitive practices committed in specific sectors. In addition, the data collected may illustrate some

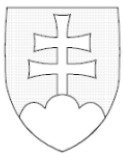
specific technical aspects of decisions or proceedings, such as the duration of the administrative proceeding, the number of pages or words. The outcome of review of the first instance decisions by the Office's appellation council or the Office´s decisions by courts, as well as the date of entry into force of such decisions, if they were upheld, may also be valuable information.

The above-stated aspects are readable from the wording and formal mandatory particulars in the text of the decision. Based on the specific textual patterns, it would be possible to identify any occurrence of these phrases in the content of the Authority's decisions in order to extract them into the metadata database relating to each Authority's decision.

According to the initial analytical phase, the Authority divided the decisions into two main groups, namely antitrust decisions and merger control decisions. Antitrust decisions primarily with the purpose of obtaining information on the application of particular legal institutes, namely leniency program and/or settlement, and merger control decisions with the intention of extracting specific information relating to the mergers under consideration.

The assignment of the Proof-of-Concept project included a methodological procedure for extracting these specific wording/textual patterns from the texts of the Authority's decisions.

Depending on the type of decision (antitrust, merger control), several specific data (attributes) were set for the extraction process.



ROZHODNUTIE

Číslo: 2021/KOH/SKO/3/56 Bratislava, 2. novembra 2021

Protimonopolný úrad Slovenskej republiky, odbor koncentrácií, v správnom konaní SK/0048/OK/2021 začatom dňa 18.10.2021 na základe oznámenia koncentrácie podľa § 9 ods. 1 zákona č. 187/2021 Z. z. o ochrane hospodárskej súťaže a o zmene a doplnení niektorých zákonov podnikateľom Jacques Bogart S.A., so sídlom 76/78 avenue des Champs-Élysées 75008 Paríž, Francúzska republika, reg. číslo 304 396 047 R.C.S. Paríž prostredníctvom jeho splnomocneného právneho zástupcu BDO Legal s.r.o., Mostová 2, 811 02 Bratislava, IČO: 51 803 330

Figure I: The number of the decision

(4) S ohľadom na skutočnosť, že podklady získané úradom v priebehu prešetrovania naznačovali, že mohlo dôjsť k porušeniu zákona, úrad začal dňa 29.09.2014 v súlade s ustanovením § 25 ods. 1 zákona konanie č. 0021/OKT/2014 vo veci možnej dohody obmedzujúcej súťaž podľa § 4 ods. 1 v spojení s § 4 ods. 4 písm. f) zákona a čl. 101 Zmluvy voči účastníkom konania PKB invest, Chemkostav a PRO-TENDER.[13]

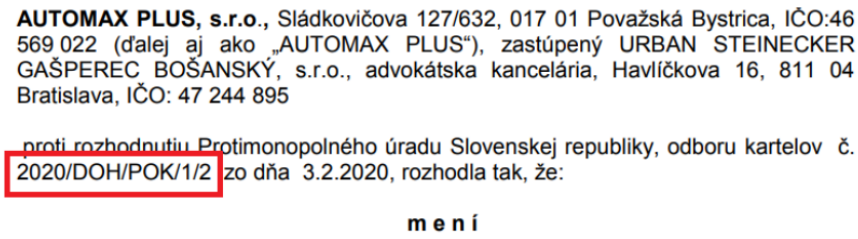
Figure II: Date of initiation of administrative proceeding

ROZHODNUTIE

Číslo: 2021/KOH/SKO/3/56 Bratislava 2. novembra 2021

Protimonopolný úrad Slovenskej republiky, odbor koncentrácií, v správnom konaní SK/0048/OK/2021 začatom dňa 18.10.2021 na základe oznámenia koncentrácie podľa

Figure III: Date of issuing decision


AUTOMAX PLUS, s.r.o., Sládkovičova 127/632, 017 01 Považská Bystrica, IČO:46 569 022 (ďalej aj ako „AUTOMAX PLUS"), zastúpený URBAN STEINECKER GAŠPEREC BOŠANSKÝ, s.r.o., advokátska kancelária, Havlíčkova 16, 811 04 Bratislava, IČO: 47 244 895

proti rozhodnutiu Protimonopolného úradu Slovenskej republiky, odboru kartelov č. 2020/DOH/POK/1/2 zo dňa 3.2.2020, rozhodla tak, že:

m e n í

Figure IV: Number of decisions reviewed by the Office's appeals council

In some cases, it was necessary to standardize and unify the search results for appropriate use in the final structured database. The wording of statements in the decisions reviewed by the Office's appellation council could have differed in case of quashing of the first instance decision; therefore, it has been suggested to standardize the database values as shown in the table below.

Typ verdiktu v rozhodnutí	Navrhovaný Typ verdiktu, ktorý bude zapísaný do databázovej tabuľky
zrušuje rozhodnutie č. 2002/PO/2/1/098 zo dňa 23.05.2002	zrušenie rozhodnutia
zrušuje rozhodnutie č. 2002/ZK/2/1/065 zo dňa 16.04.2002 a vec vracia prvostupňovému orgánu na nové prejednanie a rozhodnutie	zrušenie rozhodnutia
zrušuje rozhodnutie č. 2002/ZK/2/1/250 a vec vracia prvostupňovému orgánu na nové prejednanie a rozhodnutie	zrušenie rozhodnutia
zrušuje rozhodnutie č. 2004/KV/1/1/127 a vec vracia na nové prejednanie a rozhodnutie	zrušenie rozhodnutia

Table I: standardized database

In the Proof-of-Concept project, the Authority focused on extracting 14 key metadata (attributes) from the decisions. Some of them were only applicable to antitrust decisions or, conversely, only to merger control decisions. A number of attributes had to be further broken down, *e.g.*, parties to the administrative proceedings. With the exception of identifying data, some specific information relates to a particular entity, such as the fine imposed, other sanctions (prohibition to participate in public procurement and its duration), and the application of specific legal institutes (leniency program, settlement).

Regarding an indicative quotation for the creation of such a tool that could be used for the above data extraction, an efficiency problem has emerged. The relatively small number of decisions of the Office, around 1350, and the potential cost of the tool would be disproportionately expensive compared to manually extracting the data into a structured database, even considering the error rate of both machine and manual approaches to data extraction.

The Authority, therefore, concludes that, for the specific reasons of the smaller number of objects processed and the language specification, the introduction of similar tools is questionable for smaller competition authorities.

Slovenia
Competition Protection Agency

I. Introduction

The Slovenian Competition Protection Agency ('NCA') conducted its work until recently almost completely on paperwork. However, due to recent events and the sophisticated tools that the companies and other subjects are using to automatize their services, the NCA has undertaken some steps to narrow this gap and to introduce some tools and measures to overcome the paperwork challenges and move away from paper documentation only.

In this article, two tools will be briefly described in terms of their usage, applicability, terms and conditions of usage and the practical value that the NCA gains with the applicability of these tools. With the use of these tools, the NCA has been able to automate its processes and increase its efficiency when dealing with a large number of information, external data, and the modeling of such data.

The NCA has adapted its usual way of gathering information with current ways and contemporary measures to gather information necessary for the execution of its jurisdiction, which has led to efficiency gains and a more effective approach to analyzing of data.

II. 1KA AVK e-data gathering tool

Initially, we will focus on the e-data gathering tool that the NCA has been using for online surveys, requests for information or questionnaires that the Agency submits to different stakeholders, and subsequently the handling and processing of gathered data. The NCA opted for a source application that enables services for online surveys or questionnaires. The development takes place at the Centre for Social Informatics, at the Faculty of Social Sciences, University of Ljubljana. The University of Ljubljana is also the formal owner or addressee of the corresponding intellectual property.

It is an online service that combines support for the following functionalities:

- Development, design, and technical creation of an online questionnaire;

- The implementation of online surveys: support for invitations, publication, and distribution of data;
- Compiling and analyzing data and metadata.

The basic application can be used unlimitedly and free of charge for the purposes of online surveys for basic users, under certain terms of use.

The NCA has agreed on a special module or user upgrade of the user interface on a separate domain which allows SCPA employees to create questionnaires based on a default customized questionnaire exclusively for SCPA's needs.[41] At the same time, the module is upgraded so that the questionnaire completion certificate and the summary of answers are forwarded to both the respondent and the SCPA head office. In doing so, the SCPA satisfied the procedural requirements of the Competition Act and the General Administrative Procedure Act. SCPA pays an annual fee for the use of 1KA AVK customized module.

A. Purpose of use for the NCA

The NCA uses the 1KA AVK service for the needs of collecting information and data from a large number of addressees, as all collected answers for each individual questionnaire are kept in a single database. SCPA employees export the data from this database and then process and analyze the data locally. 1KA AVK is therefore not used for analysis but only for faster data collection, although the analysis of data collected in such a way has proved to be much easier and accessible through different software.

B. General terms

The general terms of 1KA online survey services represent a set of conditions, which users must consider when using 1KA services. As part of the General Terms, there are also Privacy policy, Cookie policy and Anti-spam policy.

C. Data Security

The communication between users and 1KA website runs via SSL standard (Secure Socket Layer) which ensures that the data cannot be accessed by third parties

[41] 1KA Enlik Anketa, https://vprasalniki.avk.si/ (last accessed June 7, 2023).

with the help of encryption. Moreover, sensitive data, which includes user passwords in 1KA database, is stored in encrypted form.

Only addressed companies can access the SCPA specific questionnaire. Each questionnaire has a different password. Recipients receive a password with a request for information (RFI) as part of a formal request for information.

D. Storage and accessibility of survey data

1KA services on servers in Slovenia permanently store and archive all surveys for users that have exclusive rights to use data collected through their surveys. A professional provider who stores the data adheres to the highest standards in terms of power, security, etc. All the data (every click of each survey separately) is stored in parallel on the secondary server (at another location). In addition, backups are performed daily on the outdoor unit, so it is theoretically almost impossible that the data would go missing.

Server log files are archived to ensure the reliability of 1KA server and they record errors in the implementation processes. Access log files are also stored and they contain time, IP address, browser type, destination file on the server and outcome of requirement (successful, failed). Log files do not allow the identification of survey responses, as they are stored separately from the user and survey data.

III. Data analyzing tool

For the purpose of more efficient analysis of different types of documents obtained during the inspections of the NCA, the NCA started to use forensic tools a few years ago when it realized that it is impossible to analyze a large number of information and data with manual tools and manual inspection of seized files. The NCA, based on the experience gathered in a number of conducted inspections, started working towards a more automated approach to the analysis of data and documents. To make the process of data gathering, analysis, and review of documents electronically accessible and eligible for an automated review using forensic IT tools, the NCA opted for a thoroughly tested IT tool called X-Ways which proved to be tailored to our needs. This tool is mainly used during inspections (on the premise inspections) and subsequent analysis of obtained documents, but occasionally also on documents that the NCA receives from the companies on request, either by RFI or by other means of information gathering (*i.e.*, during sector inquiries etc.).

A. Data security

The software provides various data protection measures to ensure the acquired data is secure and confidential. Firstly, the program creates a forensic image of the original data source in order to preserve the integrity of the data. The program also provides an option to encrypt the acquired data using AES-256 encryption, which is a widely recognized and secure encryption standard. Additionally, this software provides various authentication and access control measures to ensure that only authorized employees of the NCA can access the acquired data. These measures include password protection, access permissions and user management features. Overall, the software provides comprehensive data protection measures to ensure that acquired data remains secure and confidential throughout the investigation process. During the inspections the software is used in an isolated environment (fresh install, air-gap etc.), and in the presence of the company representatives, since the company has a right to be present during acquisition and analysis of the data. During the acquisition of the data, the NCA takes special care to the fact that only relevant data is extracted, the system is wiped clean and the acquired copies are sealed until the end of proceedings, all with the aim of ensuring the safe use of the obtained data and their adequate protection. In this phase the NCA seals the data acquired in the process of inspection and it remains sealed until it is inspected and reviewed by the NCA officials in the presence of the inspected company representatives so the transparency is safeguarded.

In the process of reviewing the documents and data acquired during the inspections, the NCA separates the relevant documentation and data eligible for the proceeding to a separate folder and hands one copy of all the relevant data to the company involved in the proceeding hence their right to transparency and exercise of their right to defense of all gathered information and data is safeguarded through this procedure. After the review process, all the data not relevant to the proceeding and not used in the process of proving an infringement is deleted from all the data storage mediums.

B. Terms and conditions

The NCA has no formal relation with providers of any software used and we acquire it on their publicly announced commercial terms. Used software has everything necessary for acquisition and analysis, such as data extraction (forensic copies or logical copies), verification, analysis (by use of search terms, or index),

deduplication, filtering, OCR, etc. Furthermore, some specialized tools are used in some instances (especially for email).

The terms of usage for X-Ways software are outlined in the EULA, which the NCA was obliged to accept prior to the usage. The EULA specifies the conditions under which the NCA can use the software, especially outlining the number of installations and users allowed, the types of activities that the software can be used for, and the restrictions on modifying or reverse-engineering the software.

The software was licensed to the NCA for use on a per-user basis, and the NCA was required to purchase a license for each installation of the software, which grants the right to use the software for the purpose of digital forensics. The NCA is required to comply with all applicable laws and regulations throughout the usage of the software, including data privacy laws and regulations. Failure to comply with the terms of the EULA can result in the revocation of the license.

The NCA does not perform any cloud analysis or other kinds of analysis where issues regarding unauthorized sharing of data to be analyzed or proprietary solutions contained within the software could arise.

C. Practical value of the software

The data analysis of this software is performed through a variety of built-in features and tools. The program uses advanced algorithms and techniques to process the acquired data and provide insights into the contents of different sources that are discovered during on-site inspections.

One of the key features of X-Ways is its ability to create a logical view of the acquired data, which allows investigators to browse and analyze the data in a hierarchical manner. The program provides advanced search capabilities, allowing the authorized employees of the NCA to search for specific keywords, file types or metadata within the acquired data.

X-Ways also allows the NCA to identify and recover deleted files and hidden data that may not be visible through traditional browsing or searching. This can be especially useful in cases where data has been intentionally or accidentally deleted or obscured which can occur even before or during investigations of the NCA.

Overall, the data analysis is performed through a combination of advanced algorithms, search capabilities, data carving and analysis tools, allowing the NCA to

gain deep insights into the contents of the acquired data. This enables the NCA to acquire evidence of different infringements during an investigation even if those stem from different digital sources. The usage of this software has widened the investigative toolbox that the NCA has at its disposal and helped the NCA to acquire, safeguard, analyze and use data regardless of a digital source and conduct a successful investigation.

Given that companies today communicate and do business in a digital environment, the use of software for analysis of vast quantities of data is a must. We use established commercial tools that are established software in the forensic environment and suits our budget and requirements.

IV. Conclusion

As set out in this article, there is an ever-increasing need to apply tools for automation that enable the NCA to perform its tasks--especially in the field of data extraction and analysis-- to be more efficient and able to tackle the challenges in the contemporary digital environments.

It has to be mentioned that the NCA does not have a designated unit or personnel dedicated only to the use of these tools. This is the reason why the NCA is facing challenges when applying these tools as it is a voluntary act of a small group of regular case handlers of the NCA that use their spare time to perform and give trainings among the group and to tackle different challenges such as the purchase of adequate equipment and hardware, adapting to different situations during inspections and performing 'mock-up' acts to gain the necessary skills and knowledge in order to be proficient in the real-time situations.

The companies under scrutiny are more and more advanced in the digital world and use a large number of computational software, therefore, the NCA is forced to follow this trend. However, due to budgetary constraints there is a 'follow up' situation happening in this area, where the NCA aspires to be the force that has the advantage in comparison to private undertakings to tackle competition issues even more effectively with the use of advanced computational tools.

Spain
National Commission for Markets and Competition

I. Short summary

This contribution was made by the Spanish National Markets and Competition Commission (CNMC) to the 2nd Annual Report of the Stanford Computational Antitrust addresses the topic of Computational tools for collusion detection.

The CNMC has among its top priorities adapting its competition enforcement to the current wave of digitization and artificial intelligence. This commitment has led to initiatives on two different grounds.

On the one hand, an active analysis and monitoring of digital business models and sectors, including the potential of algorithms and other information technologies to enable collusion and anticompetitive conduct.

On the other hand, the creation of an Economic Intelligence Unit (EIU). Among other things, this unit has the objective of applying new techniques of data analysis artificial intelligence to detect not only new forms of (algorithmic) anticompetitive conduct but also traditional ones. The unit also aims at increasing ex-officio detection and at strengthening the robustness of other investigations.

The CNMC remains committed to progressing in these two fronts in order to ensure that competition is safeguarded in digital activities and throughout the economy. However, the scope of this paper is mainly set on the use of computational tools (including, but not limited to, AI-based tools) by the EIU, which will be summarily described, considering their current stage of development.

II. Introduction

Algorithms and artificial intelligence are widely used by companies to conduct business, often with welfare enhancing effects. However, they create several challenges for competition, and hence for competition authorities.

Algorithms and other information technology tools can be used for anticompetitive purposes, by means of software tools with a varied degree of sophistication (pricing algorithms, etc.). But anticompetitive conduct can also appear

in other forms, such as self-preferencing or discrimination (*e.g.,* in rankings or matching algorithms). These risks appear more often in digital business models where algorithmic tools are used more frequently, but they can also take place in other sectors.

In order to tackle these challenges, the Spanish National Markets and Competition Commission (CNMC) has introduced digitization as a priority in its Strategic Plan,[42] leading to different initiatives in its annual Action Plans.[43] One of such initiatives is the creation of an Economic Intelligence Unit (EIU), to take advantage of new digital techniques as an aid to detect different kinds of anticompetitive conduct, including when this conduct is facilitated by algorithms and other advanced tools.

This contribution summarizes the work carried out by the Economic Intelligence Unit (EIU) in order to increase detection of anticompetitive conduct through digital tools.

III. Detection and analysis of anticompetitive conduct in the Economic Intelligence Unit

The CNMC's priorities, set in its Strategic and Action Plans, include enhancing detection capabilities and the analysis of anticompetitive conducts, among other things by refining the analysis of algorithms and artificial intelligence to gain a better understanding of their potential use in anticompetitive conduct detection.

This is one of the objectives of the Economic Intelligence Unit (EIU) of the CNMC, created in 2018. Given its tasks, the team of the EIU consists of a multidisciplinary group with economic, legal, statistical, mathematical, IT and data protection profiles, all of them with extensive experience in competition matters.

The analysis carried out by the EIU is used both for purely *ex-officio* detection and for refining the analysis in other cases: enhancing and reinforcing evidence, sampling and testing of allegations, dawn-raid planification, etc. Collecting the right data allows them to use specific techniques and analyses, such as machine-learning,

[42] CNMC, Plan estratégico, https://www.cnmc.es/sobre-la-cnmc/plan-estrategico (last accessed June 7, 2023).

[43] CNMC, Plan de Actuación, https://www.cnmc.es/sobre-la-cnmc/plan-de-actuacion (last accessed June 7, 2023).

graph analysis and other advanced techniques related to Big Data management and processing.[44]

One set of computational aids at EIU's disposal are searching tools based on an extensive data collection activity. Prior to the creation of the EIU itself, the CNMC has been working since 2015 on a public procurement database, creating specific algorithms for data extraction and cleaning, categorizing the data by quality levels regarding all participating bids and bidders.

At this point, we have to notice that, in Spain, public procurement data are stored in different repositories.

Firstly, the Public Sector Contracts Register includes basic information on the contracts awarded by all contracting bodies (central, regional, autonomous bodies, other entities governed by public law) including later amendments.

In addition to it, Public Sector Procurement Platform (PSPP) is a completely up-to-date electronic platform publishing all the calls for tenders and their outcomes. Even if according to law, all central contracting bodies are obliged to use this platform to manage their purchasing processes, regional and local contracting bodies can choose to create their own ones, provided that the platforms of the different government authorities and public entities are interconnected to establish a single platform that centralizes the publication of public sector procurement.

Then, Ministry of Finance Centralized Procurement Portal publishes calls for tenders in processes using framework agreements, centralized contracts and other centralized procedures. Likewise, this is a compulsory method for State-wide contracting bodies but optional for regional and local ones.

Finally, autonomous regions and local entities have established their own procurement data platforms.

With data coming from all these sources, EIU has hence created a complete digital database of public procurement in Spain that has already allowed the development of

[44] *See* the CNMC's contribution to the competition policy roundtable on Data Screening Tools for Competition Investigations in December 2022 to learn more about the data use and screens by the EIU. OECD, *Data Screening Tools for Competition*, https://www.oecd.org/daf/competition/data-screening-tools-for-competition-investigations.htm (last accessed June 7, 2023).

several proprietary, powerful search tools that take advantage of the processed data and currently help EIU to detect potential cases of bid-rigging.

Quite recently, the access to non-awarding data (losing bids) has opened the way to the use of machine-learning techniques that will make possible to decisively promote ex officio detection.

One of the working fields is the exploration and development and training of an AI-tool based on machine-learning algorithms, both supervised and semi-supervised, to determine the best set of features (both factual and statistically calculated) that will help the machine to build a practical model in an attempt to classify the behavior of companies participating in public procurement processes, eventually leading to the detection of collusion schemes and their classification by their specific nature or pattern, and, up to a point, the ex-ante prediction of such anticompetitive conditions on some given situations.

The second line of work in this field will be the design and building of graph databases, that store nodes and relationships instead of tables, or documents. Such databases will help in the task of analyzing relations among activity sectors, companies, and the company leaders themselves. That would hopefully provide EIU with some clustering capabilities, aiming to the prior identification of groups of companies, or even individuals with the potential of becoming a real cartel, or any other sort of collusion scheme. Both action lines are currently under intensive development, not being fully operational yet.

Other investigative approach at the EIU, different from the others but which takes advantage of some of the techniques mentioned above, is the use of Open-Source Intelligence (OSINT) and Human Intelligence (HUMINT) tools in order to provide an accurate identification and location of organizations and persons of interest, the relationships among them, and their degree of control of the companies and organizations which are under close scrutiny.

This work by the EIU is key to improve detection and investigation, and it should increase ex-officio detection of cartels, and any other anticompetitive conduct. In this regard, it is worth mentioning that between 2018 and 2022 70% of CNMC's cartel decisions have not been triggered by a leniency application. This increased rate of ex-officio cartel investigations since 2018 is partly due to the special scrutiny of bid-rigging cases (it is also important to be noted that 70% of CNMC's cartel cases concern bid-rigging).

IV. Conclusion

The enforcement of competition law in digital markets is complex, especially as much as algorithms and artificial intelligence are concerned. Competition policy tools are flexible enough to adapt to the disruption driven by digitization, but some challenges remain.

The CNMC has tried to adapt and refine its enforcement tools, for instance, through the creation of an Economic Intelligence Unit. The CNMC has also given strategic priority to monitoring anticompetitive conduct driven by algorithms and digital business models. There is a real need for advanced computational tools based on similar techniques, to ensure a more even 'playing field' among cartelists and competition authorities, and therefore some of the current efforts of CNMC are set on that target.

Taiwan
Fair Trade Commission

With rapid advancements of Internet technologies, there has been a surge of innovative business models, for example, e-commerce, cloud-computing, use of big data, the Internet of Things and sharing economies. However, the emergence of new business models may bring about further competition concerns. Taking on an important role, competition agencies need to advance their understanding of relevant markets and competitive behavior in business and equip themselves with up-to-date analytical tools and law enforcement capabilities, in order to ensure and promote market competition. In the context of increasingly complex markets, the Taiwan Fair Trade Commission (the 'TFTC') collaborated with external experts and economists to develop algorithm-based systems in 2022 that were designed to monitor major daily necessities such as soybean, wheat, corn and eggs. Moreover, the TFTC performed empirical analysis on market definition, implemented merger simulation in physical retail markets, and established geographic information systems. These 'computational tools' have been used to facilitate the TFTC's law enforcement activities.

I. Algorithm-based monitoring and detection systems

By using different methods, including Q-learning algorithms and marketing margin approaches along with web crawler tools, the TFTC developed specific mechanisms to monitor and detect whether pricing practices for soybean, corn, wheat and other daily necessities/bulk commodities were likely to raise concerns over concerted actions:

- With the Q-learning algorithm, the 2022 price and output data in relevant markets was processed and used to test and adjust model parameters. After repeating a large number of rounds and learning with the simulator, the TFTC built a simulation model of vertical and horizonal concerted actions. The analytical results showed that there was lack of evidence indicating the presence of collusion among suppliers of daily necessities/bulk commodities such as soybean, corn and wheat.
- Under the law of one price, the TFTC employed models of VAR and VECM, the Johansen cointegration test and the Granger causality test to analyze marketing margins of daily necessities. The analytical results are summarized as follows:

- Downstream prices of daily necessities including soybean, corn and wheat did not efficiently reflect their respective upstream prices. One possible explanation was that the domestic supply of daily necessities was subject to import regulations on bulk grain.
- Prices of most daily necessities remained relatively stable, but prices for some commodities, for example milk and eggs, often rose faster than they fell. More evidence was required to be carefully scrutinized to determine whether market price changes were resulting from concerted actions.

II. Empirical analysis on market definition and merger simulation in physical retail markets as well as creation of geographic information systems

As the domestic economy has undergone different developmental stages, business environments in retailing, the number of retail businesses and retail market structures have been changing. Drastic changes in market dynamics were observed in the retail sector particularly during the COVID-19 pandemic. To respond to these changes, the TFTC, external experts and economists worked together in a combined effort to develop more practical methods for merger analysis in physical retail markets. The TFTC designed and conducted consumer surveys to collect required market data and utilized economic tools to define relevant markets. The TFTC also made use of geographic information systems ('GIS') displaying the spatial distribution of retail businesses and retailer density. Furthermore, econometric methodologies were applied to emulate competitive effects in post-merger markets:

- Following the hypothetical monopoly test, the TFTC reviewed the data collected from consumer surveys to define a relevant market. The result indicated that hypermarkets across the nation could be defined as a relevant product market.
- On the basis of the 2022 price and output data in the relevant market, the TFTC made use of the logistic regression model to analyze anti-competition effects arising from a proposed merger. The model suggested that the merger was likely to have anti-competitive effects as a result of the level of the increase in post-merger prices.
- The TFTC completed domestic retail store mapping with a GIS system. The mapping result demonstrated that physical retail outlets were highly concentrated in metropolitan areas.

The TFTC employs various computational tools, including R programming, Python, GIS and digital forensic tools for the purpose of data cleaning and warehousing. In competition enforcement activities, the TFTC selects computational tools applicable and appropriate to specific cases. For example, in merger investigations, the TFTC defined relevant markets by means of econometric methodologies or GIS, and then calculated market shares and levels of market concentration. To assess unilateral effects or other anti-competitive effects, the TFTC looked at different indicators including GUPPI and compensating marginal cost reduction analysis. In respect of concerted actions, the TFTC built relevant time series models to examine trends and correlations of prices and costs, thereby identifying signs of potential concerted actions.

A. Hypermarket mergers

The TFTC made attempts with different methods to sort, warehouse and clean a large number of data sets such as tax registrations for businesses and questionnaire survey results. These methods included web crawling, text mining and machine learning in R programming and Python. Regression models and critical loss analysis were implemented to define relevant markets. The analytical process started with the hypermarket sector, i.e. the initial candidate market. The TFTC then calculated market shares and market concentration as well as assessed unilateral effects and other anti-competitive effects.

B. Cinema mergers

A new concept used by the TFTC to define local markets in cinema mergers is isochrone-based catchment areas. First, the length of time or the distance consumers would be willing to travel to reach the locations of merging parties and other cinemas was determined. On account of the diversity of domestic transport modes for personal travel--walking, bicycle, car, bus, metro and rail, the journey time or the distance was decided by the most commonly used transport modes. Isochrones were then constructed with GIS software to measure overlapping catchment areas of individual cinemas, assess the level of competition among cinemas and define geographic markets.

C. Price fixing among ice-making businesses on one offshore island

Tools for extraction of digital evidence and digital forensics were used to acquire instant messaging chats from mobile devices owned by relevant businesses. The TFTC cross-checked these communication histories and the dates of price increases and obtained direct evidence supporting price fixing.

D. Price fixing in carbonated drinks

Data warehousing and cleaning with R was applied to a large number of data sets around carbonated drinks. The impact of monthly changes was taken into consideration based on the interrupted time series approach to build a regression model for different types of carbonated drinks. The TFTC gauged the elasticity of demand by evaluating the differences in outcomes before and after price adjustments. The TFTC further compared the elasticity of demand and the extent of increases in prices to review whether price adjustments were considered economically rational.

III. In 2023, the TFTC will take the following measures to enhance the effectiveness of antitrust enforcement activities and continue technology capacity building in its law enforcement.

- Establishing applicable modules/packages for economic analysis in antitrust: The modules and packages will be tested and verified in real-life situations. The TFTC will be endeavoring to stay on top of the latest information on relevant programming languages and modules/packages for antitrust analysis to optimize the accuracy of its current models/packages.
- Strengthening expertise of digital forensics and capabilities of digital data extraction: Procuring GIS software and password cracker tools with the aim to develop cyber capabilities and reinforce the efficacy of antitrust enforcement.
- Improving mobile forensics skills and implementing new technologies in enforcement activities: To achieve efficiency improvements in antitrust enforcement, procurement of specialized tools for mobile forensics is underway. This is aimed at effectively deterring potential anti-competitive practices by collecting and analyzing relevant evidence in a faster and more precise way.

Türkiye
Turkish Competition Authority

Our mission as the Competition Authority is to prevent cartelization and monopolization, increase consumer welfare, contribute to the well-functioning of market mechanism, contribute to the improvement of international competitiveness, and to ensure that the investment environment functions in a healthy way by decreasing entry barriers.

In line with this mission, the Competition Authority has the following purposes:

- To monitor, regulate and supervise markets to prevent agreements restricting competition, abuses of dominant positions, as well as mergers and acquisitions that will significantly decrease competition;
- To promote a competition culture and make necessary disposals to ensure that public decisions and actions are made while being mindful of their impact on the competitive structure of the market;
- To conduct research related to competition law, economy, and policy, to develop policies, and to contribute to macroeconomic policies with respect to competition law.

Our vision is to be an institution which supports sustainable growth and development; takes innovation, variety, productivity, and quality to the highest level; promotes a widespread competition culture; establishes a well-functioning competitive environment; makes intellectual, economic, and administrative contributions, and is effective on international arenas.

The ongoing and completed activities for the automation of antitrust procedures and improvement of antitrust analysis in Turkish Competition Authority (TCA) are as follows:

- Various legal regulations have been made in order to increase the effectiveness of on-site inspections and to raise the standard of evidence in detecting violations by TCA. The changes made are explained below in order:

    - An update was made in the on-site inspection article, which is the 15th article of the Act No. 4054. With this change, it has become possible to examine documents in electronic media and information systems.
    - The methods by which electronic media and information systems can be examined in the on-site inspections are determined with the *Guidelines on the Examination of Digital Data during On-Site Inspections*. With this guide, uniformity has been ensured in the on-site inspections made by different teams. In addition, in accordance with the principle of transparency, it has been ensured that the undertakings are aware of the examinations made by TCA.
- In algorithm examinations, training on 'Data Science and Algorithm' was received for informing the team conducting the on-site inspection about machine learning. It became possible to make effective examinations into machine learning codes during on-site inspections after that training.
- In order to recognize new technologies to be faced in environments where inspections will be made and increase the efficiency of the inspection, work on creating know-how on the following issues are ongoing at the Authority:
    - Version control systems (Gitlab, Github, Azure, TFS, Bitbucket etc.)
    - Task management apps (Jira, Trello etc.) ;
    - In-house communication tools (Slack, Skype, Teams, Discord etc.);
    - E-mail servers (Google Workspace, Microsoft 365, Exchange vb.);
    - Cloud infrastructure systems (Cloudflare, Azure, DigitalOcean vb.);
    - Databases (MSSQL, Oracle, MYSQL, SQLite etc.) ;
    - Other (Jenkins, Elastic Search, Docker etc.).
- An Online Application System through the E-Government portal has been implemented so that applications and complaints that fall under the scope of Article 4 of the Act no. 4054 are submitted to the Authority easily and rapidly. In this way, it is possible to make a statistical evaluation of the applications submitted to the Authority.
- A competition Information Management System (RBYS) has been designed and put into practice for digitalizing, following, and keeping statistics of in-house professional processes.
- A web service application has been put into use for receiving information about procurements from the Public Procurement Authority to use such information in professional processes.
- Given the acceleration of the digital transformation, studies are carried out, especially to understand digital markets in order to adapt to changing competition rules. Some of the studies made up to date in this framework are as follows:

    - E-Marketplace Platforms Sector Inquiry;
    - Online Advertising Sector Inquiry;
    - Reflections on the Digital Transformation on Competition Law.
- R&D work for creating a data pool and detecting possible competition infringements by means of data extracting methods (scraping, crawling etc.) from web pages (social media, e-marketplace, etc.) is on-going.
- If deemed necessary within the scope of cases, examinations are made on an undertaking's IT system and security infrastructure (firewall, WAF, e-mail, log keeping and SIEM servers, etc.) to detect whether the undertaking has applied access/utilization restrictions against competing undertakings domain / IP range.
- Forensic software and hardware are used during on-site inspections in order to examine the undertaking's data stored in the digital environment. Within this framework, the most appropriate method is selected among physical imaging, logical imaging, advanced logical imaging, file system imaging, or application downgrade methods with forensic software to obtain digital data stored in personal or corporate mobile devices, which are widely used in the working processes of undertakings. In addition, system logs of mobile devices are examined, and the usage status of an application, whether it is deleted, when it is downloaded, *etc.*, are available.
- Instant messaging applications that are frequently used for communication inside or between undertakings are an important source of evidence concerning the data to be obtained during on-site inspections. Therefore, the professional staff examines instant messaging applications during on-site inspections and content, log files, and application databases.
- In order to increase the level of technical expertise in the field of forensic informatics, current developments, software and hardware updates, and new software are closely followed, and participation in forensic-related activities is ensured.
- The E-discovery feature in mail servers is used to see whether there is an e-mail deleted by the undertaking's staff in on-premises or cloud-based mail applications that keep an undertaking's e-mail infrastructure. Log records are used to detect the actions taken by the undertaking's staff on mailboxes during the inspection (deleting, sending e-mails, etc.).

United Kingdom
Competition and Markets Authority

I. Analytics and machine learning in evidence review

Throughout 2023, the CMA has continued to roll out its new evidence review platform. eDiscovery specialists and data scientists from the Data Technology and Analytics ('DaTA') unit have worked together to help the CMA's case teams exploit new analytical techniques to speed up evidence review. These include Continuous Active (Machine) Learning, email threading and clustering of similar documents. We continue to look for innovative ways to exploit our evidence review technology and, most recently, we have piloted the use of our evidence review platform to agree confidentiality representations with external parties on a cartel investigation.

The CMA is increasingly receiving submissions from parties, for merger and other cases, with documents that are not in English. While we do use professional translators, given the scale and its impact on both cost and time, we needed a solution that would allow us to quickly process thousands of documents and allow case teams to run keyword searches and review the documents as quickly as possible. Given the confidential nature of these business documents, using an external automated translation service is not appropriate. The DaTA team therefore took an open-source solution using NLP and adapted it to run on our secure data analytics platform, adding parallelism, which allowed us to auto-translate documents at scale.

II. Open-sourcing our web-scraping

Mirroring the move by business to digital, the DaTA team often finds itself scraping information from the web. The uses of scraping range from preliminary evidence collection all the way to automating remedy compliance monitoring. The CMA implements web-scaping in a careful and considerate manner, ensuring that it is proportionate, legal and does not over burden the company's servers. The DaTA unit has put together a Python package, called "selene", that standardizes our internal scraping code, by providing an (object-oriented) framework, based on open-source web-scraping packages. This approach reduces the amount of duplicated code for standard tasks from scrolling to taking screenshots. The DaTA unit has open sourced this package and publicized it widely with the hope that it can both be useful to other agencies, but also be the start of a more open and collaborative code-driven future to regulation.

III. Automating the identification of mergers and acquisitions

In the UK there is no requirement for parties to a merger to notify the CMA of a transaction. The CMA, therefore, needs to monitor news sources to identify mergers and assess whether they raise competition concerns. This has historically been a largely manual process at the CMA. The DaTA unit have recently developed a tool to track merger activity in an automated way with machine learning. The tool, which is still in development, collects news article data from various APIs and RSS feeds and passes them through a natural language machine-learning model. It then outputs a prediction as to whether each article relates to a merger. The tool should allow the CMA to dramatically reduce the amount of time it takes to sift through long lists of news articles for any sign of a relevant merger, freeing up the team to focus more on analysis of the competition concerns of each merger identified.

IV. Building our behavioral science capability

Throughout 2022, we have continued to build our Behavioral Hub within the DaTA unit. In 2023, we published a paper on experiments at the CMA which describes how and when we use field and online experiments to support our work, including in relation to online choice architecture.[45] We continue to explore and identify harm from online choice architecture, including across our portfolio of consumer protection and antitrust cases and markets work. To support this, our behavioral and data scientists have been working together to set up an eye-tracking lab to test how users interact with websites.

V. Horizon-scanning for emerging technologies

The CMA has taken steps to ensure it is proactive and forward-looking in the emergence of new technologies or emerging markets. This includes the development of its internal horizon-scanning capabilities to identify new and emerging technologies and trends in digital markets. For shortlisted technologies, our DaTA unit have produced 'technology primers' that explain the technology, the markets to which it is most relevant and the possible implications for competition and consumer protection.

---

[45] GOV.UK, *Experiments at the CMA: How and when the CMA uses field and online experiments*, https://www.gov.uk/government/publications/experiments-at-the-cma-how-and-when-the-cma-uses-field-and-online-experiments (last accessed June 7, 2023).

Large language models, or generative AI, was identified through our horizon-scanning in 2022. In May 2023, following a period of internal research, the CMA opened a review on large artificial intelligence (AI) models to understand how the rapidly evolving sector could help people, businesses, and the UK economy. The review will help the CMA develop its understanding of large AI models, known as foundation models, and their future uses in a wide range of markets as well as any competition or consumer issues that might arise.