

Unclassified

English - Or. English

8 September 2020

DIRECTORATE FOR FINANCIAL AND ENTERPRISE AFFAIRS
COMPETITION COMMITTEE

Cancels & replaces the same document of 1 September 2020

LATIN AMERICAN AND CARIBBEAN COMPETITION FORUM - Session I: Digital Evidence Gathering in Cartel Investigations

- Contribution from Brazil -

28-29 September 2020, virtual Zoom meeting

The attached document from Brazil is circulated to the Latin American and Caribbean Competition Forum FOR DISCUSSION under Session I at its forthcoming meeting to be held on 28-28 September 2020, via a virtual Zoom meeting.

Ms. Lynn Robertson, Manager Africa/MENA, LACCF ; Competition Expert - Lynn.Robertson@oecd.org.

JT03465051

Session 1: Digital Evidence Gathering in Cartel Investigations

- Contribution from Brazil -*

1. Digital evidence analysis: developing methods through cooperation¹

1. The increasing volume of digital data gathered—both in inspections and search and seizures—is a reality shared by competition authorities around the world. Despite the fact the diagnosis is uncontroversial—it is extremely difficult to extract, screen and analyse, in a timely manner, all collected data—, the strategies to deal with the issue are quite different, depending on institutional characteristics of each country, IT restraints and procedural specificities.

2. If competition authorities are facing similar challenges, there is no dispute about the fact that there is a diversity of strategies available to cope with digital evidence pervasiveness. The main goals of our contribution are (i) briefly delineating CADE's reality with respect to digital evidence analysis; (ii) explaining how and why our strategies were designed as they are; and (iii) illustrating how these strategies work using a concrete case.

3. By default, CADE only analyses digital evidence gathered from unannounced inspections such as search and seizures carried out by the competition authority itself or by the police or prosecution services, which received technical support from CADE's operations team. This allows CADE to focus on collecting evidence related only to the main targets of the investigation.

4. On the one hand, it potentially reduces the volume of digital data collected. On the other hand, this sort of strategy requires, in order for it to be successfully implemented, that an in-depth intelligence analysis of each target is carried out, otherwise it is not effective.²

5. This is the reason our current strategy was designed as it is. CADE has adopted a comprehensive approach with respect to prosecution services, the police and internal control bodies³, looking to establish formal partnerships to conduct investigations in cooperation. It has allowed our investigation team to work closely with criminal investigators to (i) gather preliminary intelligence (focusing on collecting digital evidence related only to relevant targets), and (ii) share information during the investigation, regarding relevant key-words and time lapses, contextual data and related facts (to reduce the amount of data to be analysed and expedite forensic processing).

* This contribution was written by Felipe Leitão and Valadares Roquete (Head of Intelligence Unit) from the Administrative Council for Economic Defense.

¹ This document was produced by Felipe Leitão Valadares Roquete, Head of Intelligence Unit at CADE.

² All competition authorities face the same risk: collecting only useless data (type I error) or leaving useful data behind (type II error).

³ CADE has signed Technical Cooperation Agreements with all 27 State Prosecution Services, the Federal Police and several control bodies.

6. The impact of using such strategy has been meaningful. On a bid-rigging case related to the local public transportation service around 50TB of digital evidence were gathered by prosecution services in previous search and seizures. The criminal investigation—that had begun three years prior to our own—collected relevant information about the market, the way companies behave, the constantly changes to their corporate structure and their connections with government officials. Starting forensic analyses at CADE, without accessing all the information that had already been obtained and processed would require that we process and analyse all 50TB of data, basically unnecessarily repeating what had already been done.

7. It was then that CADE decided to enter into partnership with the prosecution services using Technical Cooperation Agreement, and designed specific procedures to be used in this case: (i) first, both teams would analyse all hard-copy documents collected and share the results of their findings with each other, (ii) second, those results—including the intelligence produced by the prosecution services during the 3-year-long criminal investigation—would inform the elaboration of a search protocol, a guide for analysing digital evidence.

8. At that point, we had a list of targets (companies and their employees), solid information about meaningful dates (relevant for bid-rigging schemes) and a comprehensive narrative about the connections between companies, their employees and government officials. In the end, we were able to reduce the volume of digital evidence to be analysed from 40 to 5TB.

9. Despite the fact that we agree there is no one-size-fits-all approach to deal with digital evidence in antitrust investigations, the most important part of our strategy involves relying in our organisational and procedural skills. In a time upcoming budget restraints are almost certain, that could be an asset for competition authorities around the world.

1.1. How Project Cérebro might enhance digital evidence analysis

10. CADE's data analytics and screening project—Project Cérebro (named after the Portuguese word for brain)—was primarily designed to develop automatic investigation tools, aimed at improving on the prevention strategies used by the competition authority.

11. However, as the methods and techniques were being improved, the team responsible for the project noticed that a few tools—originally intended for investigative purposes—might be adapted to be of use to ongoing investigations. Of the many potential uses, the most promising one seemed to be digital evidence analysis.

12. Recent developments in natural language processing (NLP) and supervised machine learning (ML) algorithms opened up opportunities in these fields.

13. The team responsible for Project Cérebro is studying the advantages—as well as any possible pitfalls—of using ML algorithms to preselect digital evidence more likely to contain information relevant for the investigation.

14. Based on previous cases—in which digital evidence related to cartel practice was analysed by case-handlers—the idea is to teach the algorithm to identify digital documents with similar features, thus reducing the amount of digital data to be analysed with the forensic software. Although we most certainly are still in the very early stages of prototyping and therefore do not have any meaningful results quite yet, it does seem very promising as a path to innovation in the field of digital evidence analysis.