Organisation for Economic Co-operation and Development

DAF/COMP/WP3/WD(2022)30

Unclassified

English - Or. French

23 November 2022

DIRECTORATE FOR FINANCIAL AND ENTERPRISE AFFAIRS
COMPETITION COMMITTEE

Working Party No. 3 on Co-operation and Enforcement

Data Screening Tools for Competition Investigations – Note by France

28 November 2022

This document reproduces a written contribution from France submitted for Item 3 of the 136th OECD Working Party 3 meeting on 28 November 2022.

More documents related to this discussion can be found at www.oecd.org/daf/competition/data-screening-tools-for-competition-investigations.htm

Ms Despina PACHNOU
[Email: Despina.PACHNOU@oecd.org]

JT03508367

France

Introduction

1. While the ever increasing digitalisation of the world in which we live causes deep transformations in the economy (emergence of digital platforms, complex competitive issues to tackle, etc.), the technological innovation on which it is based can also benefit competition authorities and strengthen and improve the tools they have to carry out their missions.

2. In 2020, the *Autorité de la concurrence* (the *Autorité*) decided to set up a Digital Economy Unit within the Investigation Services to strengthen its digital resources. One of the objectives of this Unit, currently composed of four people, including two data scientists, is to develop new digital tools to help the Investigation Services in their work.

3. Among these new tools, data screening tools can provide valuable assistance to the case handlers in the detection of anticompetitive practices: cartels, price fixing and geographic market sharing are among the practices for which these tools can be very effective. These tools can also be used to automate certain market surveillance tasks.

4. The introduction of data screening tools is generally based on two distinct but interconnected components: a data collection tool and a tool for visualising data and the relevant indicators associated with it.

5. This contribution will first describe the two main methods of data collection (I), then present the main elements that the Autorité believes the visualising tool should include (II), and lastly illustrate the example of a tool currently being developed by the Digital Economy Unit within the Autorité.

1. Data collection

6. Data screening tools are generally used to monitor a market and/or try to detect anticompetitive practices (cartels, price fixing, geographic market sharing, etc.) without the knowledge of the entities concerned, at least initially, and for a certain period.

7. For these reasons, these screening tools are generally supplied by data sources that are external to the *Autorité*, and not by data collected through the *Autorité's* investigation powers.

8. There are two main options for data feeding: APIs (A) and scraping methods (B).

1.1. APIs

9. APIs (*Application Programming Interfaces*) allow access to a wide range of data within the framework of controlled access.

10. Through an API, a database owner can give fine-tuned access to third parties. It can determine what data will be accessible and who will receive the information, and can even scale access by defining different types of profiles.

11. APIs therefore allow secure and simple access to certain data. The technical architecture is taken care of by the person who makes the API available and anyone interested in the data can access it by following the protocol that has been determined. In

addition, APIs often provide access to the entire data history, which allows screening tools using this type of data to perform retrospective analyses.

12. In France, the principle of opening up public data was decided by a circular from the Prime Minister on 26 May 2011[1]; large amounts of data are available on the website "data.gouv.fr" and a set of APIs is available on the website "api.gouv.fr".

13. In conclusion, APIs are an ideal solution for long-term data screening tools. However, it is worth noting that, when using an API, the user is very often authenticated, so this solution cannot be considered if the use of the screening tool requires a certain degree of anonymity. Moreover, not all data is necessarily available via an API and in this case the scraping method can help to retrieve data accessible online.

1.2. Scraping methods

14. Scraping is a technique for the automated extraction of content from websites, or web or mobile applications, using a computer programme. The recovered data is usually then structured in a database.

15. For each use, the location of the relevant information must be identified upstream and the retrieval frequency of the desired data shall also be determined (hourly, daily, weekly, etc.).

16. The technical implementation of scraping requires a certain technicality, but many online libraries can facilitate this work. In the programming language "Python", we can mention "Scrapy" or "Beautiful soup".

17. Moreover, each scraping solution implemented is tailored according to the architecture of a page at a given time; when the page is updated, the tool may need to be readapted.

18. For all these reasons, scraping methods are not optimal for long-term projects, as the technical part requires a much higher degree of human monitoring than using an API. Lastly, this method does not allow for retrospective analyses insofar as access to the data only occurs from the moment the scraping method is implemented.

2. Visualisation and analysis tool

19. Once the data has been collected, it must be enhanced through its visualisation and analysis. A "black box" effect, where data is only analysed after the collection period is over, should be avoided at all costs.

20. The current tools allow the real-time monitoring and calculation of different indicators, which are updated every day throughout the monitoring period. The most conventional indicators that can be mentioned include the rate of price alignment (if we are looking to detect a price cartel) or geographic indicators with associated maps (if we want to highlight geographic market sharing).

21. Identifying, creating and implementing the appropriate indicators is crucial to the success of a data screening tool. These indicators are specific to each case, and therefore, a

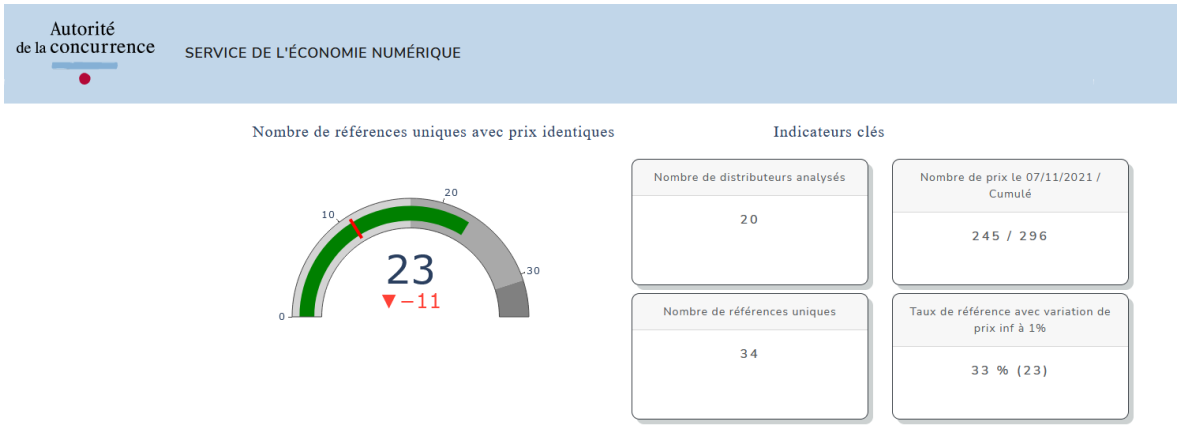
[1] Circular of 26 May 2011 on the creation of the single State public information portal "data.gouv.fr" by the "Etalab" mission and the application of the provisions governing the right to reuse public information.

good visualisation tool is set up in coordination with the case handlers in charge of the case concerned, so that they can select, create and present the most relevant indicators.

22. At the *Autorité*, the visualisation tools used are created under "Dash" (*open source Python framework*). They are presented in the form of dashboards and generally have two types of analyses, each accessible from a specific tab.

23. The first type of analysis presents general statistics and global indicators, an example of which is illustrated in Figure 1.

Fig 1: Example of general indicators from a demonstration project carried out by the Digital Economy Unit



24. The second type of analysis presents individual analyses, which are generally able to follow the evolution of data collected over time, based on narrower segmentation through a search bar. For example, it can be a page on which the evolution of product prices can be tracked over time or a page that makes it possible to find all the details of a past public procurement contract.

3. Example of data screening tools at the *Autorité*: detection of concerted bids in public procurement

25. At the end of 2020, the Digital Economy Unit launched a project to detect cartels in public procurement in France using the possibilities offered by data science. The idea at the heart of this project is to take advantage of the large amount of data available on public procurement in France and to combine it with other sources of public data, to create indicators that will allow the detection of anomalies in public procurement bids.

26. In a second phase, these anomalies will have to be analysed by the Investigation Services to confirm whether or not an anticompetitive agreement can be detected.

27. The introduction of these tools is not easy and many other competition authorities have been engaged in similar projects for several years. This is a long-term project and results are not expected for several years.

28. In France, two sources of public data on public procurement are available, namely essential public procurement data (DECP[2]) and the *Bulletin Officiel des Annonces des Marchés Publics* (BOAMP - Official Bulletin of Publication of Public Procurement Notices[3]).

29. With respect to DECP, data on public procurement contracts has been available since 1 October 2018 and represents approximately 100,000 procurement contracts per year. With regard to BOAMP, data on public procurement contracts has been available since 1 January 2015 and represents approximately 160,000 procurement contracts per year.

30. In the context of the Digital Economy Unit project, we thought it would be useful to combine the two data sources mentioned above with a third source that is complementary to the first two, i.e. the National Trade and Companies Register (RNCS[4]), which centralises all the information on companies registered in France.

31. At the current stage of the project, the connection to the above-mentioned databases is complete and operational (Figure 2). The Digital Economy Unit is now working on the design of the various relevant indicators to identify anomalous public procurement. This is a complex and often time-consuming task that requires a certain amount of ingenuity. Solutions must be found to compensate for the absence of certain information, for example in relation to losing bids. Moreover, it is not currently possible to use machine learning methods due to the lack of sufficient labelled data.

32. In view of these constraints, the project does not aim to have a high detection rate per public procurement contract (we accept a significant amount of false negatives) but hopes to compensate for this by the quantity of data available. Thus, even with a low detection rate, it is possible to detect several cases per year when this rate is applied to more than 100,000 public procurement contracts per year. In the first instance, it is therefore advisable to prioritise quantity over quality.

33. We can also hope that the detection of the first anticompetitive cartels will allow the Digital Economy Unit to progressively improve the detection tool, thus leading to the detection of more cases in the future. At least that is the desired objective.

[2] Essential public procurement data - consolidated files (DECP) - data.gouv.fr

[3] Home | Pages - boamp.fr

[4] What is the National Trade and Companies Register? 1 INPI.fr

Fig 2: Open source technical architecture project