

CONTROL OF THE COLLABORATIVE ECONOMY

IGNACIO GONZÁLEZ
GARCIA

SPANISH TAX AGENCY
(AEAT)

2017 December



IOTA

Intra-European Organisation
of Tax Administrations

CONTROL OF THE COLLABORATIVE ECONOMY

By Ignacio González García



Ignacio González García
Inspector in National Office for Investigation of
Fraud
Tax and Finance Inspection
Spanish Tax Agency (AEAT)
+ 34 917498637
ignaciomiguel.gonzalez@correo.aeat.es

Electronic commerce and collaborative economy have become realities that demand the full attention from the tax authorities in Spain. In this article it will be explained: i) The size and characteristics of this sector of the economy, ii) the strategy and tools of Tax Agency (AEAT) and, c) the use of technology for risk analysis and enforcement.

Collaborative economy

Spain is a country with more than 47 million inhabitants, and will receive during 2017, almost 75 million tourists. In what refers to e-commerce, the National Observatory of Technologies of the Information Society (ONTSI), part of the Ministry of Industry, Energy and Tourism regularly publishes a report. It has estimated a volume of 24,600 million euros in 2016 with a growth of 29,5%. AEAT has estimated that this amount is a 4,11% of total retail sales. Of the total value of e-commerce, 39,6% correspond to

purchases of goods or services carried out by Spaniards on foreign sites, 40.4% to purchases by Spaniards in Spanish stores and the remainder, 19%, to purchases from abroad in Spain, basically transport and tourist services.

In Spain the collaborative commerce is concentrated in the following sectors¹: a) lodging, representing 31.9 percent of revenue, b) transportation 27.3%, c) professional services 12.7%, d) crowdfunding 6.5%

In what refers to the lodging segment, the company with the highest volume of activity is Airbnb (USA). It offers in Spain 42,419 rental offers made by 11,795 different users. In Madrid accommodation was offered in 32,000 places providing to their owners an annual average income of €5000, which meant them a total revenue of EUR 161 million. An anonymous survey shows that 55% of owners have an income inferior to 24,000 euros. If we take into account that in Madrid there are 9,000,000 overnight stays for tourists (Airbnb is a 5% of total) and we extrapolate the figures to the total number of overnight stays in Spain we conclude that they will be about 8.93 million stays, of which their owners will obtain a revenue of EUR 449 million.

Census of e-commerce

To carry out an adequate risk control, it is essential to know the number of enterprises involved in every activity. To find out the real size of e-commerce the AEAT has used different, increasingly sophisticated procedures. Since registration for e-commerce is not mandatory, and there is no heading identifying it in any tax figure, the initial goal was to estimate the number of companies that perform it. The task was not simple because: a) a web page can be on-line without performing e-commerce automatically (without shopping cart), b) orders or reservations can be received by Internet, and it is unclear if it is electronic commerce or not. Afterwards the number was estimated using information about terminal point of sale (POS), but later, with the development of technology that we subsequently presented, it was possible to access them, count them and register them, obtaining an actual figure of more than 67,000.

INVESTIGATION TECHNOLOGY

Many tools have been created specifically for accessing Internet and analysing the information obtained. Many of them are commercial and others have been developed by tax administrations, which can be downloaded by officials. Those can be classified according to two criteria: whether personal or corporate computing tools and whether free or not.

¹ PwC report for the EU. CFR. SWD (2016) 184 final, Brussels, 2.6.2016

Capabilities that are required from these e-commerce research tools are basically of three types:

- Crawler: The term means spider. It indicates the function performed by a tool to explore the network following rules to find the information.
- Scrapper: The term indicates the functionality of a tool capacity to interpret the code in which a webpage is written and which makes it accessible by browsers such as Explorer, to find data that we request, such as a phone number or a price and copy them.
- Support tools: Databases, text editors, interfaces with corporate tools.

An administration may decide: i) to train its officers in the use of personal tools, crawlers and scrappers, so then each one of them should perform the risk analysis, crossing their data with the corporate information technology, ii) to create a unit that perform risk analysis and provides data to inspectors, iii) to create a centralized massive download tool and provide an integrated information environment in which all researchers, without the need for specialized computer skills, could perform the risk analysis. Internet data are used with a “business as usual” approach.

This third option has been adopted as AEAT's strategy. It is considered to be a very superior approach for the following reasons: a) the difficulty of training a wide number of auditors in relatively sophisticated information technologies, b) the disadvantage associated with obtaining data from different researches that could not be consolidated to obtain a global vision of the risk in the sector if they are performed with isolated personal computers, c) the loss of efficiency caused by the re-duplication of tasks and d) last but not least, obtaining the information requires in almost all cases a technical expertise, because the activity has to overcome the defenses put by the pages to avoid collapsing with requests, deal with the fact that same data has a different format in every page, and many others.

Our criterion is that the necessary skills could not be imparted in a course. It is a work of professionals. Having said that, the options, chosen are:

Scrapy as library, open source for download of data from different websites with programming in *Python*.

Scikit Learn, open source tool, programmable in Python for data analysis that allows tasks such as regression, cluster analysis, selection of models and multidimensional analysis.

Numpy. This is the package for scientific computing in Python.

SciPy. It's a library in open sources for scientific computing.

For the treatment of networks, this is for "Social Network Analysis", we can use *matplotlib*, *graphtool*, *igraph* and *networkx*.

AEAT has developed techniques to navigate on Internet with the crawler, download pages, (Scrapy), interpret them and, given that in some cases data are protected or have the format of images, process, debugg and store them in the AEAT warehousing. If we need to prompt the system to sign electronically the downloads to accredit judiciary its content, the data downloaded are identified. Downloaded data are crossed with the AEAT data stores, so that the inspector can use the data and perform a risk analysis without knowledge of these technologies.

RISK ANALYSIS AND FISCAL ENFORCEMENT

All information obtained is loaded in the warehousing of AEAT and so this type of data can be used for risk analysis and fiscal enforcement.

AEAT has downloaded using crawlers and scrappers information from: i) Webpages containing ads related to every type of hospitality, holiday residences, lodging, homestays, etc., ii) Online auctions and customer to customer webpages as Ebay or Wallapop iii) Collaborative economy, iv) platforms about restaurants as Tripadvisor v) Marketplaces as Amazon and others.

Risk analysis

- The number of tax payers under control in the area of electronic commerce is 205.991
- A total of 4.479 taxpayers are controlled by the income obtained in banners in their web pages.
- Downloaded and analysed 1.152.404 ads in the lodging sector of collaborative economy.
- AEAT has created specific datamarts for control of auction pages (168.134.986 operations downloaded).
- AEAT has: i) Identified sellers using phone number, crossing information in due form, with data provided by operators, ii) Identified sellers and premium sellers from customer to customer platforms, iii) Estimated total sales with analytics using data obtained from restaurants using, as a complement "likes" an other types of information obtained in internet, iv) Downloaded all companies that are using for a type of product Amazon as marketplace.

These data are used with several aims:

- To provide reminders on-line to taxpayers, (more than 100.000 last year) when they fill their tax declaration in the webpage of the Tax Agency, letting them to know that tax agency has specific information about an activity that is not being declared, such as a rental of a house obtained from the analysis of the ads. Taxpayers, with the knowledge that Tax Agency has information about their activity, and the surprise of the prompt that appears in the webpage, modify in most of the cases the declaration. This process has obtained impressive results.
- To use these data for analytics purposes, as a component of risk analysis, for instance in the control of restaurants.
- As a tool in the control of marketplaces.
- To provide a full vision of the economic activities of taxpayers in the collaborative sector of the economy.