Open Data Of Crime: A Review of Spanish Open Data Portals

Raquel Martin-Pozuelo
Universitat Jaume I
Spain
martinpo@uji.es

Abstract

Spain is a country with a big amount of open data in different areas like tourism, traffic or weather. The motivation of this paper is to show how difficult is for criminologist, crime analyst or researchers to find quality and detailed open data on crime in Spain and to show that for professionals is not enough to use the same available open crime data as citizens are using because the aim of data use is very different. The figure of criminologist and crime analyst is still not existing in Spain and creating them and give them a special access to police data, taking into account the data protection law, could do possible the creation of a new sector, being assessors on police offices or government in order to help to understand better the crime in a country or to create new strategies for crime prevention.

1. Importance of having open data

We cannot deny the fact of the importance of open data and its value for public policy development but also for other users. But, what is open data? Open data is a *"non-privacy-restricted and nonconfidential data which is produced with public money and is made available without any restrictions on its usage or distribution"* (Janssen et al., 2012). Data as for example, tourist information, business, traffic and weather among others are created and collected by public bodies. All these data used to be data that is "safe" to share. Other types of data that could invoke a reaction from the public is not as often shared as the others (Janssen et al., 2012).
Opening data could create a new situation where collaborative networking emerge and use and create public information (Chun et al., 2010). When the people who analyze and process data is from outside the organizational boundaries might enrich data combining with other data sources that they know and maybe even combining with data that they selves have collected using mobile phones technologies (Janssen et al., 2012).

2. Open Data in Spain

In Spain, we can find a transparency law named *"Ley 19/2013, de 9 de diciembre, de Transparencia, Acceso a la Información Pública y Buen Gobierno"* where the objective is to regulate, to spread information and to guarantee the access to that information that affect to the public organizations. There is a transparency portal[1] website owned by Government. This portal share information which

[1] http://transparencia.gob.es

knowledge could be important to guarantee the transparency provided in the transparency law. Also, it contains information requested more frequently by citizens. The public information that users can request is that data that any fellow within Public Organizations acquire in order of their job, but there are some topics with limits as for example national security, defense, public security but is possible to access to that data with partial access what means that private information must be omitted and this must not disorientate the information to keep having sense. In the case of crime data, if we read article 15 of this law we can find some rules to protect some especial data as for example personal data that could exist in requested data. In this case is required authorization from affected person. If the requested information has no private data (anonymized data), the organization could give access to this data in case that exist enough public interest on the divulgation of information. Is possible to request that data by electronic or coming to the office. The public organism where was requested that information has to reply accepting or rejecting the request within one month.

3. Open Data Crime

When we talk about open data we need to exclude private, classified or confidential data and also this kind of data that is not appropriate to publicize (Janssen et al., 2012), so what happens with crime data? Is crime data appropriate to publish? Is important that citizens know what type of crime is occurring in their surroundings? As we said above, the most data shared is that data that is safe to share, but is crime data safe to share with users and citizens? Some public servants avoid to open data because it could turn to critical questions (Janssen et al., 2012).

Focusing on the system theory states, opening data will improve governance in the way that this can have a feedback and learn more about users (Janssen et al., 2012), but for example in the Spanish police we can find some sections related on crime analysis inside police bodies but they do not share directly data. These sections are inside the national police (Cuerpo Nacional de Policía), called *"Sección de Análisis de Conducta"*[2] (Behavior analysis section) or SAC where the aim of this group is to do analysis of criminal behavior. Also, inside military police (Guardia Civil) there is a center called *"Centro de Análisis y Prospectiva"*[3] (Prospective and analysis center) or CAP where its function is to collect, spread and collaborate in topics related with security of citizens among others. Also, Guardia Civil has a Group named "*Sección de Análisis del Comportamiento Delictivo"* or SACD that create criminal profiles of youth criminal behavior, homicide, sexual aggressor, etcetera. Here we can find a good example of how the data is not opened to use or re-use by citizens. They share information that they extract themselves inside their police force but they do not share data to user's analysis.
Is good that they share data but it has not usability at all. ISO 9241-11 (1998) defined usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". They talk about "products", could we consider open data as a product? Time after, the ISO 9241-210 extended the definition from "products" to "products, systems and services" (Bevan et al., 2015). With this type of data shared we cannot achieve our goals in order to analyze data because this data is not usable. It is not in a good format to do analysis.

In other hand, we can find open data available to download in formats XLS, and CSV on the National Statistics Institute of Spain[4] (INE). In this website, is possible to select the data that we are interested on as for example, nationality and crime, type of crime, age and crime, gender and crime among others and download all these data to do our analysis, also is possible to visualize graphs of data. The origin of data depends on the type of data that we selected as for example adults with firm conviction. This data proceeds from the Central Registry of Punished, which belongs to the Ministry of Justice. Other

[2] https://www.policia.es/prensa/20110524_2.html
[3] http://gccap.bage.es/
[4] http://www.ine.es

as data of minors with firm conviction proceeds from the Central Registry of Sentences of Criminal Responsibility of Minors whose ownership corresponds to the Ministry of Justice.
Important to say that INE is not sharing where police data. They share Tribunal or Court data. Also, is important to know that INE does not share **where** the crime was committed. It shares where the crime was **judged**. The place where a crime is judged sometimes is not the same where the crime was committed, so these data shared is also not giving us information about crime in Spain.
Other website from *Ministerio del Interior*[5], we can find crime data from different police forces and it is available to download in PDF, CSV and XLS, but some of them only in PDF. This data is uploaded every 3 months but again the data offered is counties data and not as much detailed as latitude and longitude for each event.

Also, is possible to find and download data in CIS[6] (Centro de Investigaciones Sociológicas) website. They generate data creating inquest for citizens. This is not crime data from police but is very interesting for social perceptions analysis. We can find questions as *"How many crimes happened last year?"* (from the point of view if the citizens' perceptions) and we can select a second variable to add more information to our query. We can know, just asking people how many people declare suffered a crime, which type of crime and the gender of the victim. It can be very useful to extract data and to know the opinion of citizens in some topics and add that information in our research to try to understand better the society. Also, with this kind of information we can know the perception of citizens on safety/unsafety and crime. There are questions as *"Fear feeling in some situations"* in order to know where and when (day, night, at disco, in the bus…) citizens are feeling more fear. The good thing is that is possible to select by year, text free typing or advanced search. Is easy to find and easy to download data but the big issue is some of this data needs a program very complicated for user without local knowledge. Is available for downloading all the associated data as for example microdata anonymized in ASCII format in order to be used in any statistic program, the inquest they used on citizens in XLS, PDF format and data to analyze ourselves in SPSS and SAS.
Other crimes as violence against woman is easy to find, download. There is a Spanish Government website where is easy to find a Statistic portal[7] to download data related on this topic. The formats available to download the data are XLS, CSV and PDF.

We could see that finding open data in crime is, in some case, easy to download and easy to use but in other cases is difficult to download this data or just use them because we need to use software that not all citizens know how to use it. Also, is important to say that all these websites do not offer the data as a bulk but in small pieces. Is not possible to download the complete set, only is available to download per parts so that add more complexity to mix and analyze data, in other case, as we have seen, we need to make a request to the public organization in order it accepts or rejects our query, even you are not interested in personal data.

To talk about open data is important to check some details[8]. These details make interoperability data work, so data must be available to download, preferably, over the internet and in a modifiable form. It must permit re-use, redistribution and intermixing with other datasets and data must not have discrimination depending on the field of data.
Why is interoperability important? It allows to work together and to create more complex systems, to be able to combine different dataset from different sources, if they are open. That allows us to avoid having a lot of datasets but no ability to combine them to create a larger system and to obtain the real value.

[5] www.interior.gob.e
[6] http://www.cis.es
[7] http://www.violenciagenero.msssi.gob.es
[8] http://opendatahandbook.org/guide/en/what-is-open-data/

The kind of data that we can find in the previous websites is not very detailed. Is good for citizens to have some data related on crime in their city because this shows transparency to the government allowing users to compare the results obtained in our analysis using government data with the results that the government are giving to us, but if we would have access to more detailed data, a new field for researchers and also for criminologist and crime analysists could be open. These professionals could support analysis tasks from government and police departments. In nowadays these figures are not created in Spain so they cannot be hired to do these tasks. Also, we need to know if in nowadays the authority has the enough capacity to do analysis by themselves because sometimes the problem becomes from not having capacity or capability to do it but they cannot look for external professionals because the privacy and hermeticism of data.

4. Data Infrastructure in Crime Data

In Spain, we have two national police. One of them called *Cuerpo Nacional de Policía* [9] (CNP) what is civil police. The other one is *Guardia Civil*[10] what is military police. Also, we have county police as *Ertzaintza*[11], *Mossos d'Esquadra*[12], *Policía Foral de Navarra*[13] and *Policía de la Generalitat*[14] (dependent of CNP national police).

Each of these police have their own competence but they have some shared competences and the most important one is in the Spanish Constitution[15] (La Constitución Española). In article 104 explains that "*Security Forces, under the government's dependence, shall have the mission of protecting the rights and freedoms and guaranteeing citizen security".* Also, it explains that an organic law will determine the functions, basic principles of action and statutes of the Security Forces and Bodies. This organic law is "Ley Orgánica 2/1986, de 13 de marzo, de Fuerzas y Cuerpos de Seguridad"[16].
The existence of different Security Forces in a same territory and with similar functions and some of them in common[17], makes arise the need to create a law where the actuations where identic and to determinate the basic principles of acting.
Is complicated to understand how these Security Forces can work in the same territory without sharing data, even knowing that their main mission is the same.
There is no **police** data infrastructure existing in Spain and even less spatial data infrastructure (SDI). Policía Nacional and Guardia Civil have one database for International sharing finger prints[18] and also for data of dangerous missing criminals or DNA[19] of serious crimes. While these two police have direct access to these data, others as Policía Local (Local Police) should ask permissions to have access in each query. But a part of these data, they do not share daily crime data or events.

So, how are police sharing data to manage crime in each city? They are not sharing daily data. Each Security Force creates and generates their own data and they work based on what they know, but if each police body is working taking into account their own data and the information that they extract from this data, this is just a part of crime in one city and not the total amount of it, so they are not able to know what type, where or when is happening all the events or crimes, just one part of them. Also, the other

[9] https://www.policia.es
[10] http://www.guardiacivil.es
[11] https://www.ertzaintza.net
[12] mossos.gencat.cat/
[13] www.navarra.es/home_es/Temas/Seguridad/Cuerpo+de+la+Policia+Foral+de+Navarra
[14] http://www.presidencia.gva.es/web/seguridad/policia-de-la-generalitat
[15] 1978 december 29th.
[16] Orgánica, Ley. "2/1986, de 13 de marzo, de Fuerzas y Cuerpos de Seguridad." *Boletín Oficial del Estado* 13 (1986): 9604-9616.
[17] https://www.boe.es/buscar/act.php?id=BOE-A-1986-6859
[18] http://bit.ly/2rtSU99
[19] https://www.boe.es/diario_boe/txt.php?id=BOE-A-2007-17634

inconvenient is the waste of time and police officers and also duplicating data is one problem in that cases that police has the same competences. We must not forget about public organizations are the biggest data creators and data collectors (Janseen, 2011).

There is a service to centralize all emergency calls received. This service is dependent on Counties and number is 112[20], and each county has a its own service, as for example Valencian Community[21]. People can call directly to the different police departments or they can call to 112 emergency services in case of emergency. The calls arriving here are saved in another database independent of police database. This phone number is for every emergency cases as health emergency and not only related on crime emergencies.

Some organizations or governments think that opening data is will drive to a more transparent open and interactive government but one of barriers to open data is because probably if something wrong happens citizens will hold interventions and responsibility from it. So, we need to put in a balance benefits and barriers to opening data. Based on the interview done by Janseen et al. (2012) we can find two table to know what people think about benefits and barriers of opening data.

Category	Benefits
Political and social	More transparency Democratic accountability More participation and self-empowerment of citizens (users) Creation of trust in government Public engagement Scrutinization of data Equal access to data New governmental services for citizens Improvement of citizen services Improvement of citizen satisfaction Improvement of policy-making processes More visibility for the data provider Stimulation of knowledge developments Creation of new insights in the public sector New (innovative) social services
Economic	Economic growth and stimulation of competitiveness Stimulation of innovation Contribution toward the improvement of processes, products and/or services Development of new products and services Use of the wisdom of the crowds: tapping into the intelligence of the collective Creation of a new sector adding value to the economy Availability of information for investors and companies
Operational and technical	The ability to reuse data / not having to collect the same data again and counteracting

[20] http://www.112.es/
[21] http://www.112.es/comunidades/comunidad-valenciana.html

	unnecessary duplication and associated costs (also by other public institutions) Optimization of administrative processes Improvement of public policies Access to external problem-solving capacity Fair decision-making by enabling comparison Easier access to data and discovery of data Creation of new data based on combining data External quality checks of data (validation) Sustainability of data (no data loss) The ability to merge, integrate and mesh public and private data

Benefits. Extracted from Jansen et, al. 2012.

Data *per se* has not value. The data becomes valuable when it can be used. Users can use data to validate the final conclusions from data shared by government and check if these conclusions are right and justified without forgetting analysis that they can do from data to sharpen the focus of policy-making (Janssen et al., 2012). Could you imagine how important is that for criminologist or crime analyst? They could access to data, verify if conclusions are correct, give another point of view and develop strategies to address crime-related issues.

In Spain does not exist the figure of criminologist-analyst of crime. It does not exist as a police officer assistant or a department of crime assistant. So, the situation is critical for these professionals because they only can access to the same data as any citizen, and sometimes for doing some research or for creating new theories they need more or different data from the currently shared one.

Opening data has many benefits but also it has barriers. From the Janseen et al. (2012) interview, the principals' barriers that interviewee faced were at the institutional level.

Categories	**Barriers**
Institutional	Emphasis of barriers and neglect of opportunities Unclear trade-off between public values (transparency vs. privacy values) Risk-averse culture (no entrepreneurship) No uniform policy for publicizing data Making public only non-value-adding data No resources with which to publicize data (especially small agencies) Revenue system is based on creating income from data Fostering local organizations' interests at the expense of citizen interests No process for dealing with user input Debatable quality of user input
Task complexity	Lack of ability to discover the appropriate data No access to the original data (only processed data) No explanation of the meaning of data No information about the quality of the open data Apps hiding the complexity, but also potential other use of open data Duplication of data, data available in various forms or before/after processing resulting in discussions about what the source is Difficulty in searching and browsing due to no index or other means to ensure easy search for finding the right data Even if data can be found, users might not be aware of its potential uses Data formats and data sets are too complex to handle and use easily

	No tooling support or help desk Focus is on making use of single data sets, whereas the real value might come from combining various data sets Contradicting outcomes based on the use of the same data Invalid conclusions
Use and participation	No incentives for the users Public organizations do not react to user input Frustration at there being too many data initiatives No time to delve into the details or no time at all Having to pay a fee for the data Registration required before being able to download the data Unexpected escalated costs No time to make use of the open data Lack of knowledge to make use of or to make sense of data Lack of the necessary capability to use the information No statistical knowledge or understanding of the potential and limitations of statistics Thread of lawsuits or other violations
Legislation	Privacy violation Security No license for using data Limited conditions for using data Dispute and litigations Prior written permission required to gain access to and reproduce data Reuse of contracts/agreements
Information Quality	Lack of information Lack of accuracy of the information Incomplete information, only part of the total picture shown or only a certain range Obsolete and non-valid data Unclear value: information may appear to be irrelevant or benign when viewed in isolation, but when linked and analysed collectively it can result in new insights Too much information to process and not sure what to look at (Essential) Information is missing Similar data stored in different systems yields different results
Technical	Data must be in a well-defined format that is easily accessible: while the format of data is arbitrary, the format of data definitions needs to be rigorously defined Absence of standards No central portal or architecture No support for making data available Lack of meta standards No standard software for processing open data Fragmentation of software and applications Legacy systems that complicate the publicizing of data

Barriers. Extracted from Jansen et, al. 2012.

About the barriers, one of the interviewee of that analysis told that only is published data that is relatively safe. Also, sometimes are neglected the user needs for finding data and to use it. One suggestion

from Janseen et al. (2012) was to work over existing structures as a starting point in order to reinforce the actual structure because starting with a new one could be quite difficult because the current economic cuts.

Other barrier that is considerate by Janseen et al. (2012) is the technical barriers as for example the inexistence of infrastructures. As we have mentioned above, it is necessary to create a Data Infrastructure, or better a Spatial Data Infrastructure (SDI) for sharing crime data.
To create an e-infrastructure where users could decide what type of data they want to download to analyze can help to scientific researchers to obtain these data faster and better (Hey et al., 2005).
Other important thing is to focus on which open data is allowed to be published and what are expecting user from this open data (Janssen et al., 2012). Of course, the expectations about crime data from a citizen would not be the same as the expectation from a crime analyst or a criminologist. These latest two will look for data to work either in academics or as a police or government assistance. By the other hand, users usually use that data to obtain information for their selves. Is very important here to distinguish between open data access and open data (effective) use. We can compare that with the Digital Divide.

5. Difference of *use of data* and *"effectiveness" use of data*

The Digital Divide[22] is a term used to make a distinction between the gap existing between that regions that have access to information using technology (telephones, computers, television and internet) and those that do not have access or have a restricted access as for example the digital divide between cities and rural areas. As the previous one, we can talk about Data Divide and we can make a distinction between *those who have access to data which could have significance in their daily lives and those who do not* (Gurstein, Michael B. 2011). In this case, the Data Division is not as the Digital Division, because this cannot be solved just providing digital access. It is necessary that people whom now have access to these data are in the position of using this data in a beneficial way for them. So, once the data is available, who is able to make an "effective use"?
To have access to the digital infrastructure is not the same as having the skills necessaries to use data effectiveness. For example, a citizen who is looking information about his city will not need the same data as a researcher working on crime data, a criminologist or as a crime analyst could need. And, the skills necessaries to use this data will be very different.

One approach of "effective use" in open data and for that effective use, is data that could be adapted for the most number of users possible. We need to consider the language of data and technical requirements for making use if this data, but usually many of those calls for open data are from professionals whom are using data in his daily jobs, for example as research and for policy intervention purposes, and that means that usually, coming back to Digital Divide, these people who can beneficiate of this open data are the people who have resources to make effective the use of data (Gurstein, Michael B. 2011).
The principal two steps in data are access and use, but is very important to add one more step apart from these two. This is the interpretation. The interpretation is a critical analysis of data, and how and under what conditions the data available is being contextualized and meaningful, must be included. To interpret or understand data is not the same than make an effectiveness use of the data. These two are two different processes (Gurstein, Michael B. 2011). The interpretation of crime data that a citizen does could be very different of the interpretation and understanding of data that a criminologist could do. Following the article of Gurstein, there are 7 important items for effective use of open data. These include: Internet (internet access, affordability, sufficient bandwidth…), computers and software (access, time and knowledge), computer/software skills, content and formatting (available data format), interpretation/sense making (skills and knowledge), advocacy (community or individual support, availability of local resources),

[22] http://whatis.techtarget.com/definition/0,,sid9_gci214062,00.html.

governance (legal or policy regime to allow using data). As we said before is not the same citizen user or professional user so, if we focus on item "Interpretation/Sense making" we can understand that we need sufficient skills and knowledge to see what data has value and which data has not, and which data could offer us information and which data make sense to add in order to increase its value. Janseen et al. (2012) also noticed about that availability and knowledge of users to use complex data is an underestimated topic in open data and the same is happening with statistic knowledge. This is considerate for the author as a possible barrier, because opening data could contribute more to the Digital Divide focusing its use only for some groups.
As we have seen we can find not just a division in terms of data managing, knowledge and "effectiveness" usability, we also can notice that the objective to reach is not the same for user citizen than for a professional. The needs that they need are different and also the data that they are interested to request. So, should we talk about a different open data?

6. Spectrum of Data

Opening data can give us many information but there are several characteristics that we need to keep in mind depending of type of data. For example, in data related on crimes rates could give benefits but also barriers. These data could be offered in different way depending of the quality, its utility or the level of details among others. Also, this kind of data should be dependent on privacy law (Janssen et al., 2012) and personal data, as we said before, cannot be shared. In Spain, open crime data websites are not sharing all the total amount of data. These data are done and collected from government organizations which are dependent of budget from citizens. They are showing enough data for citizens that they could more or less understand what is going on in their cities, but this is not enough for professionals who have to deal with this type of data to do their work. This kind of people need more detailed data.
If we take a look on Image 1, we can observe a data spectrum where we can distinguish between Open, Shared and Closed data depending on the requirements and the kind of shared data.

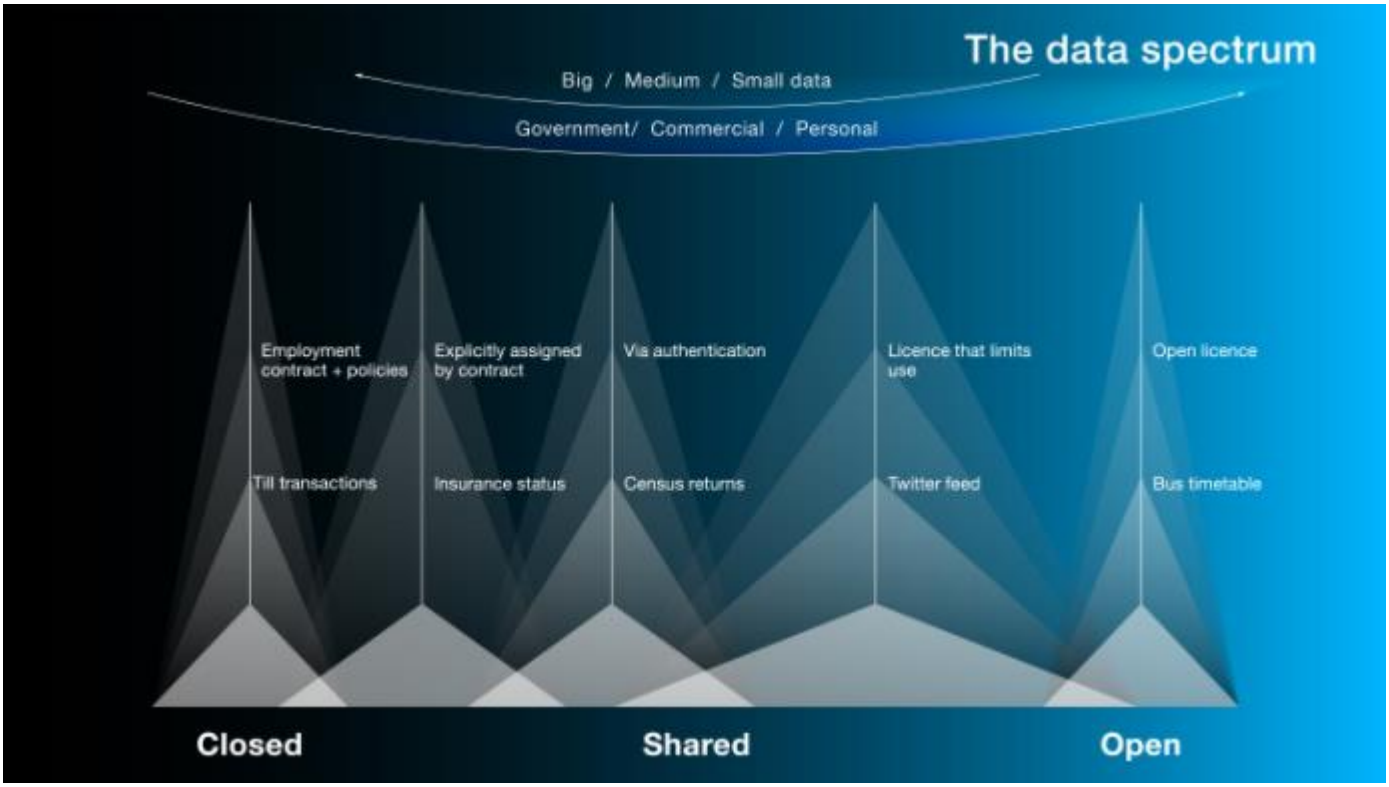

Image 1. https://theodi.org/blog/closed-shared-open-data-whats-in-a-name

For example, if we compare crime open data from Spain and data from UK we can see a very big difference. United Kingdom is sharing not only type of crime and date, they are sharing more detailed data refereeing to location. In Spain, we can download data in a global way, having spatial data as a lattice data. These data could be visualized in some software as a choropleth and we can visualize and compare per cities or counties. If we visit Data.Police.Uk[23] we can find several police department where we can download data from about crime and antisocial-behavior. The file is a ZIP where there is an Excel file and it is included the latitude and longitude where the event happened. Other website is UkCrimeStats[24] where is possible to select to download crime data by police force, by constituency, by neighborhood, by street, by subdivision, by postcode area, by postcode district, by postcode sector, by LSOA (Lower Layer Super Output Areas) this is the smaller geographical areas in England among other. You can visualize data on the website but to download the reports and the CSV file is compulsory to register and pay an amount of money monthly in most of them.

One interesting website is police.uk[25] where is possible to visualize a cluster map per zones and select the zone from where you want to download data. The data is a CSV file and contains the latitude and longitude. There is a website[26] where are available many datasets to download. This data tables are from the statistical bulletin and they are available in excel format and it shows data as a bulletin.

[23] https://data.police.uk/data/
[24] http://www.ukcrimestats.com/
[25] https://www.police.uk/
[26] https://www.ons.gov.uk/

As we can see, the difference between data from Spain and UK is huge big. In Spain is only shares, from a geographical point of view, data sets as a lattice data, showing crimes within polygons, not giving exactly information and conversely in UK the share data in a deeper way, being possible to see and analyze crime in each neighborhood, street or zip code showing crime as a cluster of points.
The difference between the two governments about what kind of data and on which detail can show it to their citizens their crime data is amazing. In any of both is existing medium term.

7. Conclusions

- Crime Open Data is existing but sometimes is hard to find, to download or to understand. Also, sometimes is hard to use if people do not have the knowledge required. Moreover, these data sometimes is only available by a formal request. Other thing to have into consideration is that from the geographical point of view (which is very interesting and very useful) some of them are sharing where the crimes were judged and not where them were committed, what could mislead to users of the data.
- It would be good if the police forces share the events and crimes in the same data base in order to avoid the duplicated data and to be more conscientious about what is happening in every space. Could be useful to create a Spatial Data Infrastructure.
- We must fight to have a portal where will be easy to find all data crime from **police** forces and in the same site, data from Government, Tribunal, etc. We need some data without filters (excluding personal or protected data) to compare the conclusions and results shared from other institutions or organizations.
- A good improvement could be taking into account the data spectrum to give access to external professionals to more detailed information as, and for example, more geospatial data for that professionals that should work with this kind of data. Also, this could be a very good opportunity to create the figure of the criminologist and crime analyst as an assistance of government, court, police forces.

References

Bevan, Nigel, James Carter, and Susan Harker. "ISO 9241-11 revised: What have we learnt about usability since 1998?." *International Conference on Human-Computer Interaction*. Springer International Publishing, 2015.

Chun, Soon Ae, et al. "Government 2.0: Making connections between citizens, data and government." *Information Polity* 15.1 (2010): 1.

Gurstein, Michael B. "Open data: Empowering the empowered or effective data use for everyone?." *First Monday* 16.2 (2011).

Hey, Tony, and Anne E. Trefethen. "Cyberinfrastructure for e-Science." *Science* 308.5723 (2005): 817-821.

Janssen, Katleen. "The influence of the PSI directive on open government data: An overview of recent developments." *Government Information Quarterly* 28.4 (2011): 446-456.

Janssen, Marijn, Yannis Charalabidis, and Anneke Zuiderwijk. "Benefits, adoption barriers and myths of open data and open government." *Information systems management* 29.4 (2012): 258-268.

Standard, I. "Ergonomic requirements for office work with visual display terminals (vdts)–part 11: Guidance on usability. ISO Standard 9241-11: 1998." *International Organization for Standardization* (1998).