A Guiding Architecture and System for Making Crime Data Available in Brazil

Edvan Soares
Federal Rural University of Pernambuco
Zip Code: 52171-900 - Recife/PE
Brazil
edvan.soares@ufrpe.br

Vanilson Burégio
Federal Rural University of Pernambuco
Zip Code: 52171-900 - Recife/PE
Brazil
vanilson.buregio@ufrpe.br

Antonio Jorge Delgado
Federal Rural University of Pernambuco
Zip Code: 52171-900 - Recife/PE
Brazil
antonio.delgado@ufrpe.br

Kellyton Brito
Federal Rural University of Pernambuco
Zip Code: 52171-900 - Recife/PE
Brazil
kellyton.brito@ufrpe.br

ABSTRACT

Brazilian Public Security Departments disclose crime data from their respective states in order to maintain transparency and openness. In fact, the opening of these data not only increases the society's awareness of the problem but also gives to it the opportunity to be more participatory. However, despite the evident benefits and effort to make this data available, there is still some issues on how such data is visualized and presented via Brazilian official portals. Most of the time, crime data is formatted in tables without any statistical information about the facts, making it difficult to have a precise overview about how crimes take place. This paper presents an architecture and system model to improve the availability and visualization of crime data in Brazil with the aim of providing a better visualization experience for those who access this information, allowing them to identify crimes hot spots as well as relevant patterns and trends.

CCS CONCEPTS

• **Information systems** → Mashups; • **Software and its engineering** → Software architectures;

KEYWORDS

Data crime, Public security, Brazil crime, Opening data

dg.o '17, June 07-09, 2017, Staten Island, NY, USA

ACM ISBN 978-1-4503-5317-5/17/06…$15.00
http://dx.doi.org/10.1145/3085228.3085312

ACM Reference format:
Edvan Soares, Vanilson Burégio, Antonio Jorge Delgado and Kellyton Brito. 2017. A Guiding Architecture and System for Making Crime Data Available in Brazil. In *Proceedings of dg.o '17, Staten Island, NY, USA, June 07-09*, 2017, 6 pages.
DOI: http://dx.doi.org/10.1145/3085228.3085312

1 INTRODUCTION

It is well known that the opening of government data can bring great benefits to society. Increases in transparency and democracy control, popular participation, citizen empowerment, innovation, improved effectiveness and efficiency of government services as well as measurement of the impact of public policies are among the main benefits [1].

In Brazil, for example, several initiatives have been taken to ensure that more government data be open to the public. Among the main initiatives is the Brazilian Open Data Portal (Portal Brasileiro de Dados Abertos - http://dados.gov.br), which is a federal government initiative that provides different types of data such as health data, transport system data, education indicators, public security data, and so on. In addition to the publication of data, many events such as hackathons and application contests are held to encourage citizens to use open data in a way that promote better services for society [2]. Still in Brazil, besides the federal government, many states and cities have their open data portals and promote various other incentives for the proper use of their available data.

Despite all the effort to make data usable, many issues can still be found and need to be overcame. For example, in the State of Pernambuco in Brazil, the Department of Social Defense monthly publishes in its portal the data of lethal crimes occurrences, which is the kind of data this paper deals with. These data are available in several separated PDF files, which contains a list of murders occurred in each month. It is possible to find information such as victim's name, date of occurrence, sex, age and others. Nevertheless, the main problem lies in how the data is presented, especially because of the format in which they are made available. It is not possible, for example, neither to obtain clear and objective statistics nor filter specific information. In addition, this format does not attend all the

principles of open government data, mainly the one about being "Machine processable" which establishes that data must be reasonably structured to allow automated processing [3].

With the availability of crime data, it is possible to obtain benefits such as: (1) Minimizing the workload of police; (2) Increase community awareness about crimes, which provides more cooperation; (3) Assist in community policing and problem solving; (4) Increase public awareness of neighborhood issues; (5) Facilitate partnerships with researchers and other departments; (6) Prevent police agencies from being misinterpreted; and (7) Increase accountability of law enforcement agencies [4]. However, to obtain these promising benefits offered by the provision of data, it is necessary to provide such data in a way that society can understand and use them.

In this context, this paper aims to present a system model and architecture for crime data, where users can view the data more clearly as well as use them to create several other initiatives. The remainder of this paper is organized as follows: Section 2 presents the background on public security challenges and issues of data on public security. Section 3 presents our mapping approach of violent crimes, including proposed architecture, technologies and frameworks, and implementation. Finally, Section 4 presents concluding remarks.

2 BACKGROUND

2.1 Public Security Challenges

Violence is an old problem and has probably always been part of society, where its practice can be related to various factors such as attitudes and behaviors, social, economic, political and cultural conditions. Many governments have looked for a variety of solutions to try to reduce violence, since in addition to the high cost produced in the public coffers there is also the cost of pain and suffering that are caused in the victims of this problem [5].

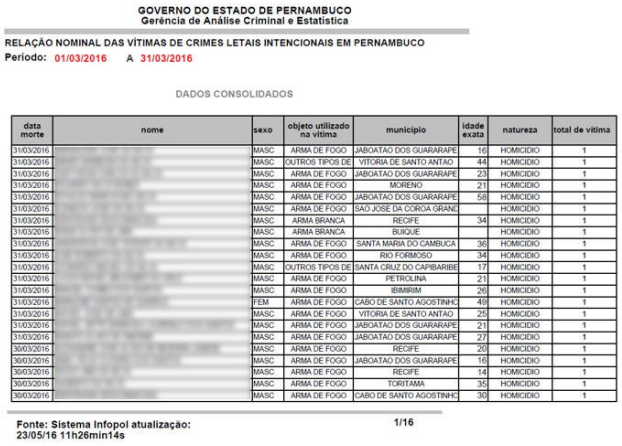
In a survey conducted in Brazil [6], violence appears as the third problem that most worries Brazilians, where 57% of interviewed people considered that. The same survey, points out drugs issues as the second-place concern, being considered by 61% of the interviewees, and in the first place is corruption which was considered by 65% of the interviewees. It is possible to realize that although violence is the third-ranked problem, first- and second-ranked problems are factors that strongly influence the increase of violence, which makes this data even more worrisome.

In this sense, the Brazilian government has created several initiatives to solve the problem of violence. One of these initiatives is the dissemination of crime data with the aims of raising awareness of the problem and increasing interactions between governments and citizens.

2.2 Issues of Data on Public Security

In Brazil, states have a direct responsibility for the protection of citizens through public security services, where each one has its military police – in charge of ostensive and uniformed patrolling - and the civil police – in charge of crimes investigations [7]. Both military and civil policies are subordinated to the public security department of their respective state, which is responsible for coordinating and drawing up guidelines for the police's work, as well as promoting clarity and transparency for society about police occurrences. Thus, periodically, the state's public security departments publish crime data in their portals to keep the citizens aware about violence-related data.

In practice, there is no pattern followed by the states in publishing their security data; each state often adopts different publishing models. In the state of Pernambuco, for example, the public security department provides a monthly list of some information on crimes, as per illustrated in Figure 1.



GOVERNO DO ESTADO DE PERNAMBUCO
Gerência de Análise Criminal e Estatística

RELAÇÃO NOMINAL DAS VÍTIMAS DE CRIMES LETAIS INTENCIONAIS EM PERNAMBUCO
Período: 01/03/2016 A 31/03/2016

DADOS CONSOLIDADOS

data morte	nome	sexo	objeto utilizado na vítima	município	idade exata	natureza	total de vítima
31/03/2016		MASC	ARMA DE FOGO	JABOATAO DOS GUARARAPE	16	HOMICIDIO	1
31/03/2016		MASC	OUTROS TIPOS DE	VITORIA DE SANTO ANTAO	44	HOMICIDIO	1
31/03/2016		MASC	ARMA DE FOGO	JABOATAO DOS GUARARAPE	23	HOMICIDIO	1
31/03/2016		MASC	ARMA DE FOGO	MORENO	21	HOMICIDIO	1
31/03/2016		MASC	ARMA DE FOGO	JABOATAO DOS GUARARAPE	58	HOMICIDIO	1
31/03/2016		MASC	ARMA DE FOGO	SAO JOSE DA COROA GRAND		HOMICIDIO	1
31/03/2016		MASC	ARMA BRANCA	RECIFE	34	HOMICIDIO	1
31/03/2016		MASC	ARMA BRANCA	BUIQUE		HOMICIDIO	1
31/03/2016		MASC	ARMA DE FOGO	SANTA MARIA DO CAMBUCA	36	HOMICIDIO	1
31/03/2016		MASC	ARMA DE FOGO	RIO FORMOSO	34	HOMICIDIO	1
31/03/2016		MASC	OUTROS TIPOS DE	SANTA CRUZ DO CAPIBARIBE	17	HOMICIDIO	1
31/03/2016		MASC	ARMA DE FOGO	PETROLINA	21	HOMICIDIO	1
31/03/2016		MASC	ARMA DE FOGO	IBIMIRIM	26	HOMICIDIO	1
31/03/2016		FEM	ARMA DE FOGO	CABO DE SANTO AGOSTINHO	49	HOMICIDIO	1
31/03/2016		MASC	ARMA DE FOGO	VITORIA DE SANTO ANTAO	25	HOMICIDIO	1
31/03/2016		MASC	ARMA DE FOGO	JABOATAO DOS GUARARAPE	21	HOMICIDIO	1
31/03/2016		MASC	ARMA DE FOGO	JABOATAO DOS GUARARAPE	27	HOMICIDIO	1
30/03/2016		MASC	ARMA DE FOGO	RECIFE	20	HOMICIDIO	1
30/03/2016		MASC	ARMA DE FOGO	JABOATAO DOS GUARARAPE	16	HOMICIDIO	1
30/03/2016		MASC	ARMA DE FOGO	RECIFE	14	HOMICIDIO	1
30/03/2016		MASC	ARMA DE FOGO	TORITAMA	35	HOMICIDIO	1
30/03/2016		MASC	ARMA DE FOGO	CABO DE SANTO AGOSTINHO	30	HOMICIDIO	1

Fonte: Sistema Infopol atualização: 23/05/16 11h26min14s

1/16

Figure 1: Data of lethal crimes in Pernambuco, Brazil.

In the portal of the Department of Public Security of the state of São Paulo, the publication is made differently from the state of Pernambuco. Lethal crime data are published annually in a table where there is only information of crime category and number of occurrences per month, as shown in Figure 2.

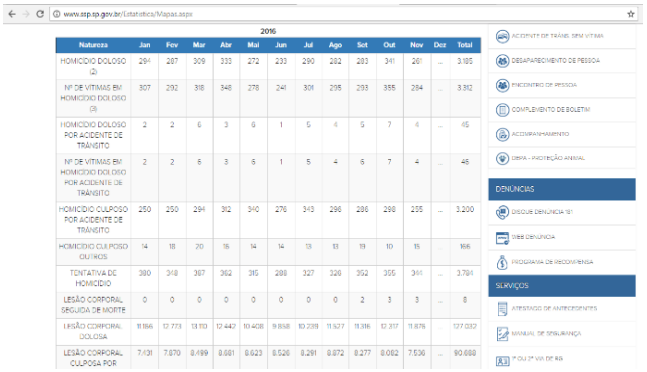


Figure 2: Data of lethal crimes in São Paulo, Brazil.

The same lack of standards for data publishing is found in other public security portals provided by the other Brazilian states, since they also significantly varies the way data are made available, what leads to unclarity and confusion for those who seek it. Considering the crime data of the state of Pernambuco as a reference, it is possible to realize some problems with respect to the format in which data is made available. Firstly, it is the fact that data is shown only in lists within PDF files. Although the data has some information such as sex, age, city, and others, there is no possibility of customized filtering, like listing the crimes by the sex of the victim or by city. Another problem is

that for a typical user, who does not have more advanced knowledge of technology, it would be impractical to generate statistics from this information since data is embedded in PDF files and cannot be processed automatically unless there is a prior treatment of them. On the other hand, in the data provided by the state of São Paulo, the main problem is the lack of details. It is possible to filter information by county, month and type of crime, however, there is no important information such as the date of each crime specifically, the gender and age of the victim, among others details that are relevant to those who search for crime data. Another difficulty is the fact that the data is published annually, thus, only at the end of the year people have access to such information.

Although each portal has some specific problems, a common problem among Brazilian portals that provide crime data is the lack of a unified format that could be easily reused by citizens interested in creating new initiatives to turn raw data more useful and beneficial to society. Another common problem among portals is the lack of clarity in data visualization, where understanding of information is somewhat hampered. In this sense, the proposed architecture and system presented in this work aims to solve the problem of visualizing crime-related information, giving more clarity and freedom to those who seek it, as well as creating a model of data availability that allows people to use available data to create new tools that benefit society.

3 A MAPPING APPROACH OF VIOLENT CRIMES

In the model development we used the data of lethal crimes of the state of Pernambuco, Brazil. This choice was based on the fact that the Pernambuco public security department provides more specific data, whereas many other states provide only the most basic ones. Thus, the system model for mapping lethal crimes would have a more complete set of information, serving as an example for other states.

3.1 Crime Statistics

In the development of the statistical data of the project, in addition to the data crimes, the system uses statistics from the Brazilian Institute of Geography and Statistics (IBGE), which provides relevant information such as the number of inhabitants in a city, number of women and men, among other information that help enrich data presentation and evaluation.

Among the services provided, it worth highlighting the one that presents homicide rates of a certain city along the time. This rate considers the number of deaths per 100,000 inhabitants and encompasses homicide crimes in their different variations such as: intentional homicide, injury resulting in death and robbery and murder. To obtain this rate, the following calculation was used:

$$\frac{number\ of\ intentional\ lethal\ violent\ deaths}{number\ of\ inhabitants\ of\ the\ city}\ X\ 100{,}000$$

Another type of statistics that is generated by the system is that of crimes by sex. For this statistic the occurrences are separated by year, where the user can check how many female and male people were homicide victims in a given year. Another type of statistics are cases of homicides separated by age group, where users can obtain information on the number of victims of this type of violence in a given year, where the age ranges are separated by ten-year intervals. It also shows statistics based on the victim's skin color, in this statistic it is possible to obtain information on how many people of a given color were victims of homicide in a given year. In the extracted data, the color information was divided into the following categories: brown, black, white, yellow and uninformed. Finally, another statistic that is presented is an annual ranking of violence among the cities of the state, where it begins with the city that has the highest rate of violence and ends in the city that has the lowest index. To obtain this index, the same calculation already presented in this subsection is used.

The data used to generate these statistics were the data that are made available by the Department of Social Defense of the State of Pernambuco, which is limited to the information that was presented. Therefore, it is important to emphasize that this model of statistical presentation is not unique, and should be increased with other data that will enrich the information.

3.2 Architecture

Inspired by the social machine paradigm [8] and idea of deriving government as a social machine [9], the developed architecture consists of an online platform that makes use of different data sources of interest to seek crime-related data and some statistics provided by official Brazilian institutions. An overview of the proposed four-layered software architecture is illustrated in Figure 3. Each layer is briefly described next from bottom to top.

Data Source Layer. In the bottom layer, there is a set of different sources. Such data sources include crime-related data (e.g., police-reported crimes in pdf files), social indicators and population census (e.g., official statistics provided by IBGE).

Data Processing Layer. On top of the data layer, a set of repositories is responsible for wrapping specific data sources. Each repository has a set of specialized data-processing components. Such components are in charge of reading, filtering and extracting information of interest prior to storing it in a data store for future analysis and integration. The Data Analytics Controllers provide a unified access point for managing the data gathered by the different repositories. Further, it also provides additional analytics capabilities to support the service layer.

Service Layer. This layer uses the Data Analytics Controllers to provide several services, such as crime managements and filtering services and statistical timeline analysis. These services are exposed as Application Programming Interfaces (APIs) and are grouped into two categories, namely i) *Crime Basic API* and ii) *Crime Statistics API.*

i. *Crime Basic API* – these services refer to the key management, filtering and grouping operations on crime data, which are useful to build different views of data, such as crimes by gender, age, weapon type and many others not represented in Figure 3.
ii. *Crime Statistics API* – these services are based on a set of data analysis tasks made on crime data in conjunction

with population census and social indicators to derive different rates, trends and insights. Such services include statistical timeline analysis, city violence ranking, homicide rates and others.

Application Layer. Finally, this layer is in charge of supporting the required applications that can be built on top of the online platform. These applications can combine different platform's services to enable the development of Web-based tools, mobile, apps, mapping solutions, statistics dashboards, just to mention few.

3.3 Technologies and Frameworks

In the construction of the first part of the project, a backend with REST services was developed using Spring Boot, which supports the development with the Spring framework (https://projects.spring.io/spring-framework/) that uses Java as programming language. It is worth noting that since the purpose of this work is fulfilled, there is no objection in using other technologies. In this specific case, Spring Boot was chosen because it is an alternative to the Spring MVC, which is a very consolidated framework, but its configuration is not so easy. Spring Boot has the Spring MVC maturity, but with an important improvement that is the abstraction of the complex Spring MVC configuration processes, which greatly streamlined the development process. The backend developed with Spring Boot is directly connected to a MySQL database where crime and population statistics information are stored. Thus, along with Spring Boot, Hibernate (http://hibernate.org/) is used as a persistence framework. Hibernate aims to simplify the process of working with database, since many things are done in an automated way, such as creating tables, inserting, removing, updating, and reading data. Further, Hibernate supports the creation of queries through a powerful query language called Hibernate Query Language (HQL) which is an object-oriented query language that greatly facilitates and reduces the working time of handling data.

In the second part of the project some online tools were developed in the application layer. These tools receive data from the backend in JavaScript Object Notation (JSON) format and presents them in a clear and organized way. The built Web-based tools use Bootstrap (http://getbootstrap.com/), which is a front-end framework that uses HTML, JavasScript and CSS to support responsive design in a way that created web pages can adapt themselves to any screen size/resolution. The framework AngularJS (https://angularjs.org/) was used to manipulate the data in the user interface. AngularJS uses JavaScript as programming language to provide a "Two-way data binding" feature, which allows templates to be rendered in HTML according to the data contained in a pre-defined scope (model), so any change made in the view is reflected to the model and vice-versa, without the need to create additional code for manipulating the Document Object Model (DOM) directly. In practice, AngularJS is responsible for dealing with HTTP requests to the backend server and updating the data models automatically in the user interface (aka frontend). Along with AngularJS, some complementary plugins are used to render data into charts, such as Angular Chart (http://jtblin.github.io/angular-chart.js/) and the Angular Morris (https://angular-morris.io/). Finally, keshif (https://keshif.me/) is used to create a Web-based exploratory dashboard that helps users to interact with crime data, customize comparisons and discover unexpected trends.

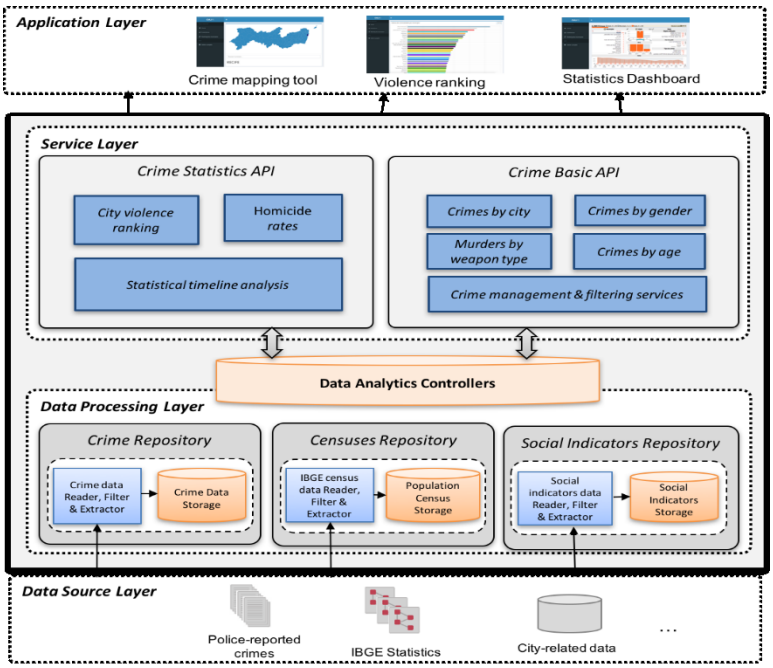


Figure 3: Four-layered Architecture for making crime data available

3.4 Outline Implementation

The implementation was done in a way to keep the system intuitive by making information available in a clear way and easily understandable by users. The system's homepage (Figure 4) presents a map of the Pernambuco state divided by cities on which the user can click and information about the crimes of the selected city is shown. In addition, a search field is also offered so that the user can select a city by typing its name. A dashboard with different statistical data charts is also implemented, as per illustrated in Figure 5. These charts aim to convey information quickly and clearly, and are targeted at any audience, from expert to non-statistic expert users interested in following and understanding violence trends in their hometown. These charts are customizable and we use colors and types that make the information more suggestive for those who are comparing data between different cities, for instance.

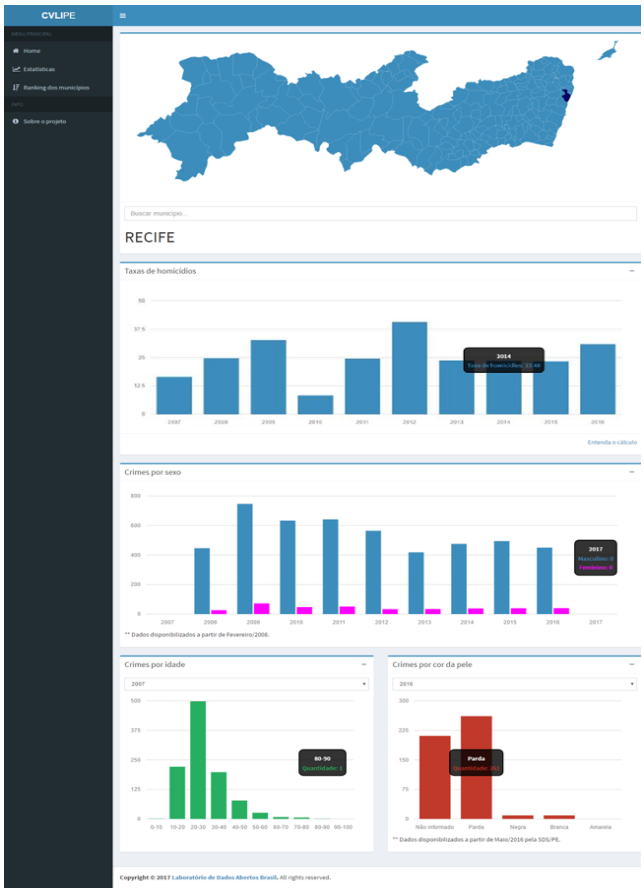


Figure 4: Initial screen of the system.

Deeper insights into data statistics is easily obtained from making comparisons between data from different perspectives. Summarizing and filtering data according to specific interests are also possible such as, for example, searching for crimes that took place in a time interval defined by the user, filtering homicides by the type of object used to kill the victims or even any combination of these filters and other filters. Journalists, police officers, researchers and many other professionals can benefit from the dashboard provided by the tool.

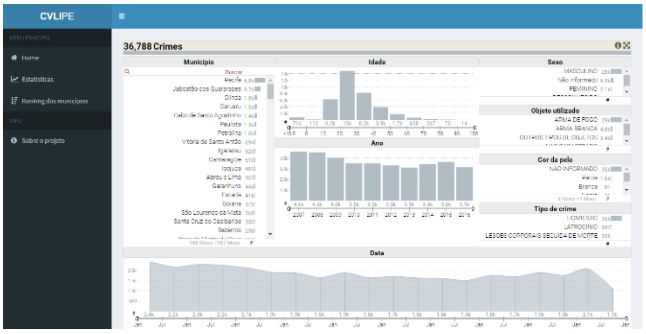


Figure 5: Web-based exploratory dashboard

Finally, Figure 6 shows the violence ranking screen which presents a list of cities ranked by their rates of lethal crimes, in which is possible to check changes over the past ten years. This visualization helps citizens to make comparisons between crime rates of different cities, and may help the society to increase its critical sense and awareness of how effective is the public security provided by its governors.

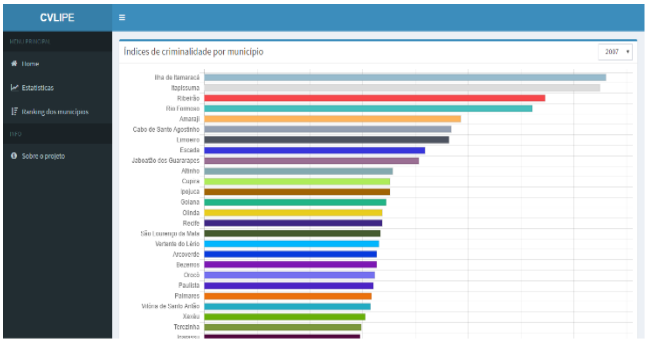


Figure 6: Violence ranking of cities based on rate of lethal crimes.

4 CONCLUSION

The Brazilian government has been seeing the awareness of people through the dissemination of crime data as an alternative in the fight against violence. However, in the way that the data are disclosed there is great difficulty in understanding this information, since in most cases the way of presenting this data is not intuitive and offer little or no statistical information. Aiming at this problem, this paper presents a guiding architecture and a system that besides proposing a better experience in the visualization of this information, makes this data available in a format that is understandable by machines so that other people can create different initiatives on top of the proposed online platform. Since the model was created using real data, the creation of statistics was limited to the data available. This points to the importance of the release of more specific data by the government, such as the complete address of the event, the time, among other data that could further enrich the available statistics. As we provide online open APIs, we believe

that the developed system will naturally conduct to the establishment of a plethora of crime-related applications and services, built by developers with a great interest in a more effective public security service. In more than one sense, the proposed architecture can offer several possibilities that converge to the fully participation of everyday people in a multitude of current and future violence prevention initiatives. Future work includes the development of some participatory-democracy services to support cyber participation of citizens against crimes, as well as the integration of violence data with other sources of open government data that will help to make this system model more comprehensive under other social and economic aspects.

ACKNOWLEDGMENTS

This work is partially supported by INES, grants CNPq/465614/2014-0 and FACEPE/APQ/0388-1.03/14.

REFERENCES

[1] Open Data Handbook. Open Data Guide. Available in: <http://opendatahandbook.org/guide/en/>. Accessed on January 10, 2017.

[2] dos Santos Brito, K., Costa, M.A.S., Garcia, V.C. and Meira, S.R. de L. (2014). Experiences Integrating Heterogeneous Government Open Data Sources to Deliver Services and Promote Transparency in Brazil. *IEEE 38th Annual Computer Software and Applications Conference*. Vasteras: IEEE, 606-607.

[3] The 8 Principles of Open Government Data. Available in: <https://opengovdata.org/>. Accessed on January 10, 2017.

[4] Wartell, J. and McEwen, J.T. (2001). Privacy in the Information Age: A Guide For Sharing Crime Maps and Spatial Data, pp. 54. Institute for Law and Justice, NCJ 188739, Washington DC.

[5] Dahlberg LL, Krug EG (2002). Violence – a global public health problem. In: Krug EG et al., eds. World report on violence and health, pp. 3–21. Geneva, World Health Organization.

[6] RSB 28 - Problemas e Prioridades. Corrupção é o principal problema do país. Available in: <http://www.portaldaindustria.com.br/cni/publicacoes-e-estatisticas/estatisticas/2016/01/1,80708/rsb-28-problemas-e-prioridades.html>. Accessed on January 10, 2017.

[7] Portal Brasil. Segurança Pública é dever do estado. Available in: <http://www.brasil.gov.br/governo/2014/10/seguranca-publica-e-dever-de-estado>. Accessed on January 19, 2017.

[8] Vanilson Burégio, Silvio Meira, and Nelson Rosa. 2013. Social machines: a unified paradigm to describe social web-oriented systems. In Proceedings of the 22nd International Conference on World Wide Web (WWW '13 Companion). ACM, New York, NY, USA, 885-890. DOI: http://dx.doi.org/10.1145/2487788.2488074

[9] Vanilson Burégio, Kellyton Brito, Nelson Rosa, Misael Neto, Vinícius Garcia, and Silvio Meira. 2015. Towards Government as a Social Machine. In Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion). ACM, New York, NY, USA, 1131-1136. DOI: http://dx.doi.org/10.1145/2740908.2743976