

Geoprocessing Applied to Open Government Data

Daniel Farias Batista Leite, Julio Henrique Rocha,
Claudio de Souza Baptista, Ana Gabrielle Ramos
Falcão

Federal University of Campina Grande, Campina Grande,
Paraíba, Brazil
{danielfarias, juliorocha}@copin.ufcg.edu.br,
baptista@dsc.ufcg.edu.br anagabriellee@gmail.com

Hugo Feitosa de Figuerêdo

Federal Institute of Education, Science and Technology of
Paraíba, Monteiro, Brazil
hugo.figueiredo@ifpb.edu.br

Abstract—The increasing social supervision in various sectors of society, particularly in the government administration, has demanded for information access policies in several countries. In Brazil, for instance, there is a law that regulates access to the Open Government Data (OGD) for reading, monitoring and reusing data in new projects and applications. In this scenario, several applications that extract information from open data are being developed throughout the world. This paper presents the *BrasilMaps* system, an open source application that integrates, through GIS technology, twelve open data government services offered to the Brazilian citizens. Data as Basic Health Units (BHU), Notary Offices, National Employment System (SINE), Service Unit of the Inland Revenue, Universities and Federal Institutes, Agencies of the Ministry of Labor and Employment and the Federal Police Offices are made available in the proposed system. The proposed system provides information that helps people in the use of public services. Through the access to this data, citizens can benefit from the nearest services offered, identify the quality of some services and filter the desired services.

Keywords- Open Government Data; Geoprocessing; GIS.

I. INTRODUCTION

Over the last years, geoprocessing technologies have played an important role in the provision of solutions for the public sector. GIS presents a great potential for information dissemination, reflecting in several sectors of the society, especially in the public sector, in which we can note an intense social control. The use of GIS enhanced with social media enables a higher supervision and participation of the population in government policies.

As a consequence of such social control in the public sector, many countries have invested in information access policies, such as the United States and New Zealand. In Brazil, in particular, two of these laws stand out: the Complementary Law 131/2009, which states the transparency of the financial execution of the Union/State/City; and the Public Information Access Law 12.587/2011.

The Public Information Access Law (IALaw) establishes the access to Open Government Data (OGD), which are data produced by the government and made available for the citizens in order to enable not only its monitoring and accessing, but also its reuse in new projects, websites and

applications. Furthermore, the government data are only considered open when published in conformity to eight principles: complete, primary, timely, accessible, machine processable, non-discriminatory, non-proprietary and license-free [1]. In addition there are three important laws [2]:

- whether the data cannot be found or indexed on the web, it does not exist;
- if it is not open and available in machine readable format, it cannot be reused; and
- if any legal framework does not allow it to be repurposed, it is not useful.

The adoption of Open Government Data policies is beneficial in many ways. At first, with the transparency of the government actions, one can create a collaborative social control environment for both the population and the government. Matheus et al. [3] also highlight another benefit of the adoption of Open Data, which is the possibility of generating new information and applications from the interaction between the government and the society through the exploration of open data derived from several data sources.

In this paper, we propose an GIS application named *BrasilMaps*[4], that enables the integration of heterogeneous data sources from twelve data bases of public equipments (extracted from the Open Data National Portal - "Portal Nacional de Dados Abertos"), using geoprocessing techniques and allowing visualization and searching on such data.

The database contains data on Basic Health Units, notary offices, Employment National System, Internal Revenue Service, *Fundacentro* (responsible for the registration of work accidents and diseases), Universities and Federal Institutes, Ministry of Labor and Employment and Federal Police offices. It also contains data about services provided for assisting individuals in threat or right violation situations (physical, psychological or sexual violence, amongst others), such as the Social Assistance Specialized Reference Center the Social Assistance Reference Center, Therapeutic Communities (specialized in sheltering people with disorders derived from psychoactive substances) and Social Assistance Private Networks.

In order to provide a good visualization on the map of the aforementioned sources, it is necessary to use a methodology based on the Extract, Transform and Load (ETL) process.

Therefore, the data sources will be extracted and examined so that only the ones that qualify are loaded and, if necessary, normalized. In many data sources, we can find incomplete and uncertain information due to several factors that may occur during the data collection.

When using many data sources, some challenges may rise on the data treatment process, notably: normalization of the geographic information, information filtering, elimination of imprecise information, correction of uncertainties and errors, incomplete information and outdated data. Such challenges were overcome using a methodology for the data treatment obtained from the aforementioned sources, thus facilitating the exhibition of the information in the map.

The rest of this paper is organized as follows. In Section II, we discuss related work and applications that make use of Open Government Data. Next, we address the methodology used for obtaining and manipulating the open data in Section III. In Section IV, we focus on the *BrasilMaps* application in details, including its architecture and available functionalities. In Section V, we describe the usability assessment performed to evaluate *BrasilMaps*, and finally, in section VI, we present our conclusions and the future work.

II. RELATED WORK

It is understandable that, due to the short lifetime of the IALaw, the Open Government Data policy in Brazil is still under the process of development and improvement. In this sense, few applications like *BrasilMaps* were been developed. Mootiro Maps [5] works with five public services datasets and shows all the information in only one layer at map, so it is impossible to filter information. Moreover, a single icon is used to represent all the public services. PAC II [6] displays the location of public constructions but does not distinguish the icon according to the category of construction, only a single icon is used.

Pedroso et al. [7] highlight several problems that some datasets made available by the government agencies, such as their availability in only the .pdf format, which makes interoperability difficult to achieve. In this perspective, Lourenço and Serra [8] defined a maturity model composed of 5 levels, in which the level 0 indicates the available databases that do not present any structure, whereas the highest maturity level (level 4) indicates that the available information is structured.

Breitman et al. [9] performed a quantitative and qualitative analysis of the datasets available in the Brazilian Open Data National Portal, indicating the categories the datasets are organized and the total volume of each category, the origin of each dataset (spreadsheets, web services, among others) and the format they are made available (XML, Json, among others). On the other hand, Xu and Zheng [10] performed an analysis on the open data available on the world under the user perspective, indicating which categories the population would like to have access to and how such data should be made available.

Matheus et al. [3] and Correa et al. [11] performed a comparative study between several Brazilian transparency portals, regarding the federal, state and city levels, coming

to the conclusion that, none of the portals are in conformity to the eight aforementioned principles for data to be considered open. Furthermore, Erickson et al. [12] argue that applications developed using open data should be regulated according to some rules, such as the way the data must be explored.

Even in such unfavorable scenario, however, many applications using open data information has been developed in Brazil. Nonetheless, these related applications use only a single database and, most of them, do not explore the spatial dimension of data. They only display the information in the form of charts and tables. Artigas [13] and Artigas and Chun [14] highlight the importance of exploring other means for displaying data, especially geolocation. The exploration of the spatial dimension may enable the detection of tendency and patterns between datasets. Graves and Hendler [15] state that displaying the data in many forms is essential for the entertainment of the citizens. In this context, this paper presents the *BrasilMaps* application that apart from integrating data from twelve open databases, enables the spatial visualization and interpretation of the data.

III. DATA TRANSFORMATION PROCESS

The *BrasilMaps* application performs the spatial integration of twelve databases. However, in order to accomplish the integration of these heterogeneous databases a data transformation process was necessary, once some databases did not have geolocation information of the underlying equipments. Figure 1 illustrates the data transformation process implemented.

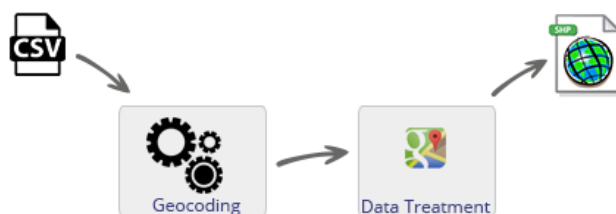


Figure 1. Extract, Transform and Load (ETL) Process.

All the files were collected using the comma-separated values format (.csv), thus being in conformity to the technical rules for the publication of Open Data, other than preserving the original structure of the information. Then, these files were submitted to the geocoding module. The geocoding module is responsible for performing an analysis on the geographical information of the databases and converting them into geographic coordinates, in case they are in the form of addresses. This conversion is called Geocoding, which is the process of identifying the geographic coordinates (e.g., latitude and longitude) from other geographic information, such as an address or a postal code. In *BrasilMaps*, the geocoding process is performed by a web service available from *Google Maps*.

It is known that it is very unlikely to achieve a full conversion of the address information into geographic coordinates. This is due to many reasons, such as the absence of complementary information (postal code, neighborhood), the misspelling of street names, and others; and all this data that were not successfully geocoded were discarded. It is better not to display them than displaying them erroneously. At the end of the refining process, about 92% of the submitted data were successfully converted. The failures encountered happened due to: incomplete address information of the public service; typos or misspellings from the team responsible for gathering the open government data; and outdated addresses. At the geocoding stage, we must mention the high success rates obtained for the data related to SINE and *Fundacentro*. Figure 2 presents the success and fail rates for each database considered, making it possible to note the low success rates of the other databases, such as the Federal Revenue information: only 77%.

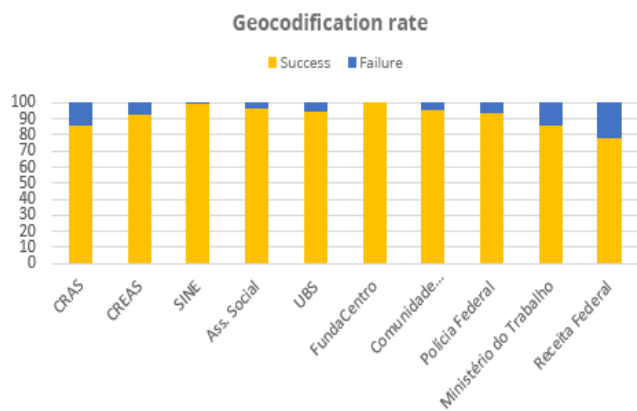


Figure 2. Geocoding success rates.

After the refining process, in which all databases presented associated geographic coordinates (latitude and longitude), it is necessary to convert them into spatial objects compatible with the spatial Database Management System chosen: PostgreSQL/PostGIS. The spatial data conversion module is responsible for this task, performing the instantiation of a new spatial object through PostGIS's *ST_Geom_FromText* function.

Finally, after the data transformation process we used Java Database Connectivity (JDBC) for sending SQL instructions to the database, loading the spatial database with the data gathered, considering each database as an entity responsible for storing equipment specific information. Next, we used *OpenLayers* to import and publish the entities with the support of the map server *GeoServer*.

IV. BRASILMAPS

This section introduces *BrasilMaps*, an application that allows the visualization, in the spatial dimension, of the government services offered to the population. This system was developed using the programming language *JavaScript* and *HTML5*, thus allowing its access via web navigators, even from mobile devices. *BrasilMaps* was awarded with the

first prize in the 2nd Brazilian National Open Data Competition.

The main goal of *BrasilMaps* is to provide a Geographic Information System (GIS) that enables the loading, processing and displaying of the Open Data for the population (made available from the Federal Government), offering a simple and practical public service search mechanism in a region of interest. The system's ability to categorize each public service results in a potentially higher and more effective participation of the society in consuming the open information that until then was of hard access to.

BrasilMaps makes use of the user's current location to display the public interest services locations closer to where he is or to any other given location. User location is detected from the Internet once it has been authorized by the user, as Figure 3 shows. When the user accesses the system from a Global Positioning System (GPS) enabled mobile device, a higher precision on the detection of the current location is achieved. This way, when the access to *BrasilMaps* is done in a big city (São Paulo/SP, for instance), the accurate location detection assists the user when performing searches and also results in faster responses from the system, once only the public service within the user's region of interest will be shown.

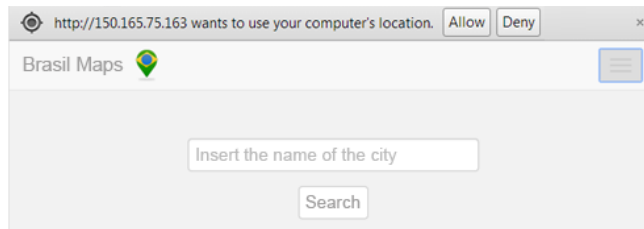


Figure 3. *BrasilMaps* initial page.

Figure 4 illustrates the moment of the user's geographic location detection when accessing *BrasilMaps* from a mobile device.

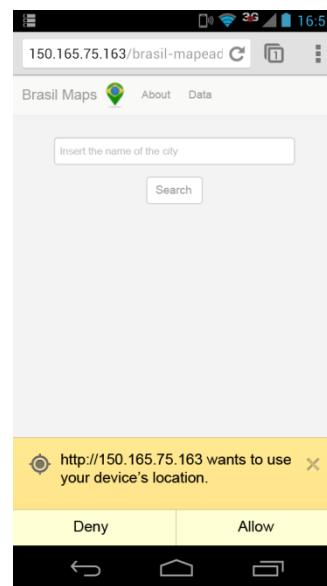


Figure 4. Accessing *BrasilMaps* from a smartphone.

In both situations, in case the user chooses not to share his location, a search field is provided for him to insert any address or city desired. This way, the application is launched with the map centered on the location submitted.

The following subsections describe **BrasilMaps** architecture and its functionalities details.

A. Architecture

BrasilMaps was designed according to the 3-tier model, aiming to facilitate portability, remote data updating and system modularization. The three layers that composed the system are: the Application Layer, the Business Layer and the Data Layer. Figure 5 illustrates the system's 3-tier architecture.

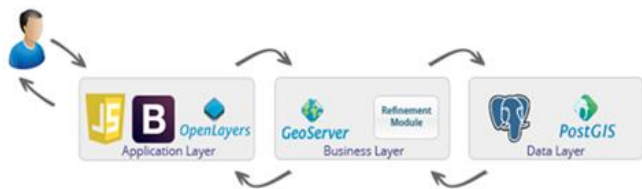


Figure 5. **BrasilMaps** architecture.

The Application Layer, responsible for presenting the information for the final user, provides mechanisms for loading, managing and viewing the data in a simple and intuitive mode. This layer has all the functionalities necessary for the users to benefit from the information of their interest. The Application Layer is responsible for communicating with the Business Layer performing requests, handling responses and formatting data properly to correctly display the information to end users.

The Business Layer is the intermediate element between the Application Layer and the Data Layer. This layer is responsible for managing the flow of information consumed

and forwarded by the application in terms of requests and responses.

The Business Layer encompasses the *GeoServer* open source map server framework. *Geoserver* provides *Webmapping* development services. From a connection with the spatial database, *GeoServer* can produce a set of thematic maps which, in turn, will be consumed by the Application Layer from a *Web Map Service* (WMS) request. The geocoding module, that contains the geocoding service responsible for obtaining the geographic coordinates, is also located in the Business Layer.

Finally, the Data Layer is composed of an object relational database with support to spatial information (PostgreSQL + PostGIS). This layer is responsible for persisting and managing all the application's data, dealing with SQL queries and spatial functions.

B. Functionalities

The spatial location of all twelve public equipments available in the application needs to be presented in a map. In this context, **BrasilMaps** contains a map viewer (*WebMapping*), with the zooming and panning functionalities. The map displayed on the application was built using the *OpenLayers* library, that provides the visualization of spatial data.

Each public equipment is displayed in the map as a layer. This way, **BrasilMaps** enables overlaying the information, in which different equipment layers may be combined in a map. Figure 6 presents **BrasilMaps** main interface, which is launched with the map centered on the user's location, with the layer "Notary Offices" and "Basic Health Units" enabled. On the right side of the image is shown the *LayerSwitcher* component, responsible for managing the layers.

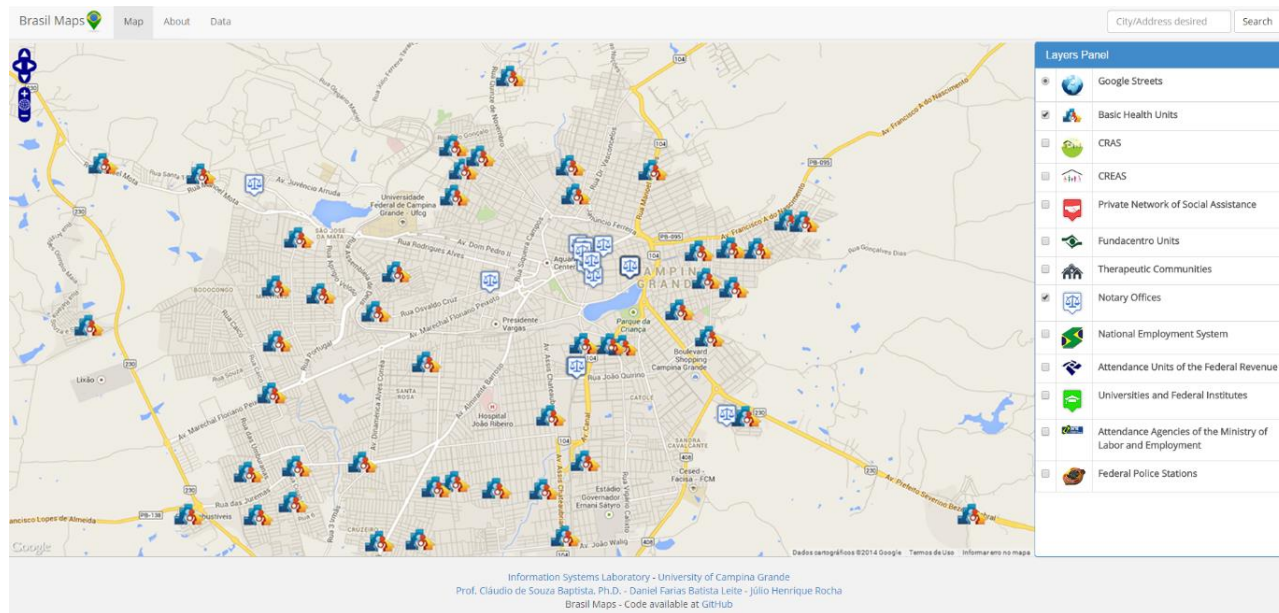


Figure 6. **BrasilMaps** main interface with two layers enabled.

Through user's location *BrasilMaps* assists the user in finding services of his interest. For example, let us consider a user that has been in an accident at his home and needs fast medical care. Hence, *BrasilMaps* can, using user location, point the Basic Health Units closer to him, facilitating the search for medical help and probably avoiding greater damages due to delayed medical support.

It is vital for a system that embraces and displays public services to provide the details for each of these services. Thus, *BrasilMaps* allows users to view the details of each entity displayed on the map using an information window. To use this functionality, the user must click on the desired public equipment, which will cause a window to popup displaying all the basic and specific information of the equipment, such as phone number, physical structure and others. This capability enables the refinement of the desired searches. In the example given above, the user might analyze whether the chosen Basic Health Unit would meet his needs. Figure 7 presents the information window of a Basic Health Unit.

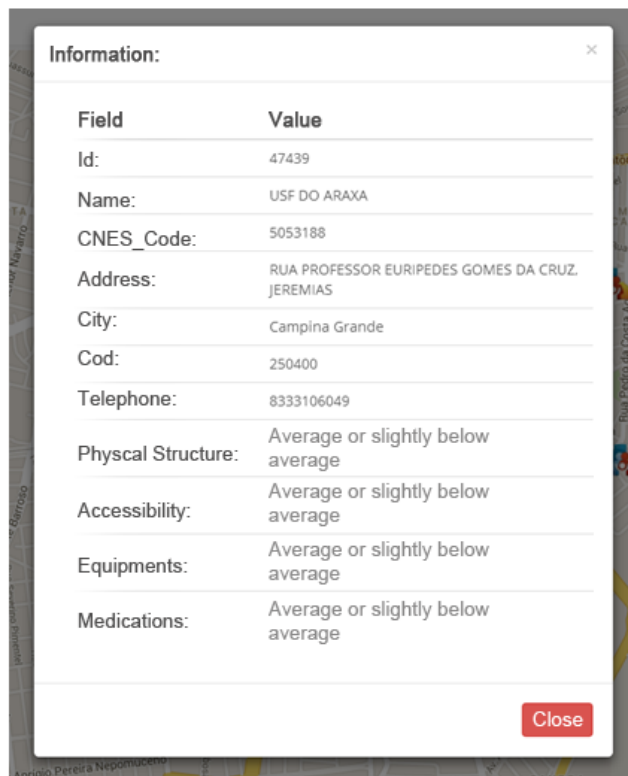


Figure 7. Popup with the details of a Basic Health Unit

With the help of the information window, a user that wishes to register a recently acquired property can check if the closest notary office to his current location may perform such task, as shown in Figure 8.



Figure 8. Popup with the details of a notary office.

Informing the user about the specific services offered by government services close to him promotes knowledge and avoids unnecessary displacement to obtain such information.

V. USABILITY ASSESSMENT

In order to assess the usability of the proposed system, we performed an evaluation consisting of analyzing its conformity to the standards defined in chapter 14 of ISO 9241 - the standard from the International Organization for Standardization that covers ergonomics of human-computer interaction. The recommendations described in such chapter of ISO 9241 suggest how the user interaction with the system might occur, considering specifically the menu structure and presentation, navigation, option selection and execution. So, such recommendations were evaluated in *BrasilMaps* and checked if they were successfully adopted. The results are shown in Table 1.

TABLE 1. RESULTS OF THE STANDARD CONFORMITY EVALUATION

Standard	#Sar	#Ar	AR (%)
ISO 9241 Part 14	31	37	83,78

Where #Sar is the number of successfully adhered recommendations, #Ar is the number of applicable recommendations and AR is the Adherence Rate ($AR = (Sar / Ar) * 100$).

Therefore, we can observe that the adherence rate is over 80% for the analyzed chapter, meaning that *BrasilMaps* is in very acceptable conditions of use.

VI. CONCLUSIONS AND FUTURE WORK

This paper presented an overview of *BrasilMaps*, a GIS application that aims to assist the citizens in finding the desired government services closer to their location. As future work, we plan to incorporate crowdsourcing techniques combined with VGI (Voluntary Geographic Information) for feeding *BrasilMaps* with information obtained from social networks.

With the goal of expanding the user experience with the system, in the future *BrasilMaps* will support spatial searches operations and will allow the generation and visualization of routes between the user location and the desired public service. Lastly, the integration with other government databases, whether national or international, is of high importance. However, the adoption of international databases will create new system evolution possibilities and challenges that will be explored in future works.

REFERENCES

- [1] The Annotated 8 Principles of Open Government Data, available at <http://opengovdata.org/> [accessed: 2015-01-05].
- [2] The Three Laws of Open Government Data, available at <http://eaves.ca/2009/09/30/three-law-of-open-government-data/> [accessed: 2015-01-05].
- [3] R. Matheus, M. M. Ribeiro, and J. C. Vaz, "New perspectives for electronic government in Brazil: the adoption of open government data in national and subnational governments of Brazil", in Proc. of the 6th International Conference on Theory and Practice of Electronic Governance - ICEGOV, October 2012, pp.22-29.
- [4] BrasilMaps, available at <http://150.165.75.163/brasil-mapeado> [accessed: 2014-10-25].
- [5] Mootiro Maps, available at <http://maps.mootiro.org/project/307/map> [accessed: 2014-12-21]
- [6] PAC II, available at <http://pac-info.herokuapp.com/map.html> [accessed: 2014-12-21]
- [7] L. Pedroso, A. Tanaka, and C. Cappelli, "The Brazilian law of access to information and the technological challenges of open government data", Brazilian Symposium of Information System, May 2013, pp. 523 – 528.
- [8] R. P. Lourenço and I. Serra, "An Online Transparency for Accountability Maturity Model", in Proc. of the 13th IFIP WG 8.5 International Conference- EGOV, September 2014, pp. 35-46.
- [9] K. Breitman et al. "Open Government Data in Brazil", IEEE Intelligent Systems, vol. 27, no. 3, November 2012, pp. 45-49.
- [10] H. Xu and I. Zheng, "Open government data: from users' perspectives", in Proc. of the 7th International Conference on Theory and Practice of Electronic Governance - ICEGOV, October 2013, pp. 366-67.
- [11] A. S. Correa, P. L. Correa, and F. S. C. Silva, "Transparency Portals versus Open Government Data. An Assessment of Openness in Brazilian Municipalities", in Proc. of the 15th Annual International Conference on Digital Government Research – DG.O, June 2014, pp. 178-85.
- [12] J. S. Erickson, A. Viswanathan, J. Shinavier, Yongmei Shi and J. A. Hendler, "Open Government Data: A Data Analytics Approach", IEEE Intelligent Systems, vol. 28, no. 5, March 2013, pp. 19-23.
- [13] F. Artigas, "Spatial analytics for open government data", in Proc. of the 15th Annual International Conference on Digital Government Research – DG.O, June 2014, pp. 357-58.
- [14] F. Artigas and S. A. Chun, "Visual analytics for open government data", in Proc. of the 14th Annual International Conference on Digital Government Research – DG.O, June 2013, pp. 298-99.
- [15] A. Graves and J. Hendler, "Visualization tools for open government data", in Proc. of the 14th Annual International Conference on Digital Government Research – DG.O, June 2013, pp. 136-45.