Caring for My Neighborhood: a platform for public oversight

Gisele S. Craveiro and Andrés M. R. Martano

University of São Paulo
Av. Arlindo Béttio, 1000 - 03828-000 - São Paulo - SP - Brazil
{giselesc,andres.martano}@usp.br

Abstract. Social participation is one of the strong claims that have been done about benefits derived from open government data, but to achieve this goal there are many social, technical and congnitive barriers to discuss. Regarding the specific example of budget transparency, despite there is data supply on governments portals, it is not understandable yet for a broader audience. In order to address this challenge we present in this paper Cuidando do Meu Bairro (Caring for My Neighborhood), a tool that was adopted in São Paulo city to promote citizen engagement and better visualization of public budget expenditures. From unstructured and semi structured information about public spending, some expenditures are geocoded and exhibited on the city map. The color code used in their pins reflects the real time spending status, which delivers budgetary content in a more accessible form to the public. We also discuss some challenges faced, the initial users demands and others ideas to discuss in our ongoing work, the Project Cuidando do Meu Bairro. A broader picture of this project is presented in order to give an idea of potential for linking government information about budget actions, budget amendments from the municipal legislative, and the citizen participation in the budgetary process.

Keywords: geocoding public budget, data visualization, open government data

1 Introduction

Public budget should express an action plan in order to meet all the needs and priorities of the people. Its implementation should be as transparent as possible to allow extensive monitoring by the society. Since 2000, Brazil has been improving its accountability and transparency with the Fiscal Responsibility Law [1], amended in 2009 by the Supplemental Law 131 [3], also known as the Transparency Law. This law states that all Brazilian public entities have to web publish detailed budget data in a 24-hour basis.

Since 2008 the open data movement introduces new elements related to the way public data is accessed, used and reused [13]. Open government data (OGD) refers to releasing freely accessible, standardized and easily readable data. It

promises to make governments more transparent, accountable and efficient. It is claimed that it can also foster greater civic participation and promote new business opportunities. Governments, entrepreneurs and civil society organizations all over the world are interested in exploring the potential of open data. As a result, several governments have set up open data portals releasing budget and other public datasets on the Internet.

As the Transparency Law is previous to this, there is no mention about machine readability or other OGD principles [2]. The only guideline closest to them in the legal framework is that all Brazilian public entities must to provide budget data in downloadable datasets, but OGD has been receiving growing attention and some transparency web portals now are compliant to some OGD principles [10]. Government has released hundreds datasets building the basis for several stakeholders to use and re-use information. The idea is to set up collaborative space to partner with civil society and civic entrepreneurs in the development of public services. Individuals and civil society organizations have been developing digital applications in order to generate many points of view regarding these data [9].

Besides the data supply, it is also important to address the point of readability and accessibility of open data, as public information is the foundation of the participatory open data eco-system. But only public information is not enough to create participation, it is key to understand the demand, and the growing importance of this citizenship that is empowered by these data. This can be particularly complex, as we can see in the open budget data case. Open budget data understandability has, at least, two barriers: proficiency in the use of technological tools and knowledge about the public accountancy domain.

Also regarding the demand side for budget data, civil society generally is more focused on the subnational level (cities, districts, counties) budget plan and its execution monitoring. This can be explained by several factors that range from the local and more focused interest (to include and monitor specific projects that benefit a particular community) to a wider scope (national public policies monitoring and evaluation). The scenario of growing decentralization is particularly important in Brazil, where the municipal level plays a fundamental role to deliver health and education public services.

Concerning these motivations we have developed a project named "Cuidando do Meu Bairro"[1] (Caring for My Neighborhood, henceforth CMN). This project aims to provide a tool for citizens to exert social control and oversee individual expenses in the public facilities of their cities. In order to achieve this, São Paulo City's public spendings are geocoded and displayed on a map, allowing anyone interested to follow, in a real time way, individual expenditures. This tool aims to contribute bridging the supply and demand sides.

The remainder of the paper is structured as follows. Related projects are presented and briefly discussed in Section 2. Section 3 presents the tool, its advances, difficulties and other demands. Section 4 addresses our ongoing project and other possible extensions and finally, Section 5 concludes this work.

[1] http://cuidando.org.br

2 Related Work

The International Aid Transparency Initiative (IATI) which is a multi-stakeholder initiative that seeks to improve the transparency of aid, developed an XML standard used to share detail on aid projects. The standard is documented[2] and data is available[3]. The standard includes a *Transaction* element[4] which can contain detailed information on financial flows, and also a *Location* element[5] which now makes use of the Mapping for Results geocoding methodology to represent location information. The tools are available in the platform AidData[6]. Development Tracker[7] is another platform that has been developed by UK's Department for International Development (DfID) on top of some IATI files from the government and its partners.

Another important initiative to analyse and visualize public spending is developed by the World Bank in the BOOST program[8]. It seeks to enhance accessibility and use of fiscal data for enhanced expenditure analysis as an input to improved budget processes and outcomes. The geographical tagging is done in a more aggregated level in order to highlight public policies implementation's performance instead of tracking individual projects or activities. The same idea is found in the visualizations delivered by the Open Knowledge's OpenSpending tool[9].

In scientific literature we have found the discussion about budget geocoding in [14]. This paper examines how the display of municipal budget data via web mapping technology allows citizens to visualize how the budget affects their lives and neighborhood. In this study, citizens access Cincinatti's municipal capital budget data via a budget mapping tool. A marker or pin on the map was generated for each capital budget project listed in the database. Markers were color-coded depending on their category. The information provided was not on a real time basis, but the conclusions show that displaying data in map form decreases the barriers citizens go through in order to process and understand budget data. It also increases relevance of the data by showing citizens the effect of the budget in their neighborhood.

3 CMN Tool

The CMN tool essentially collects semi structured public data (XLS files) daily disclosed by the Planning Secretary of São Paulo City's website[10], and places

[2] http://iatistandard.org
[3] http://iatiregistry.org
[4] http://iatistandard.org/activities-standard/transaction
[5] http://iatistandard.org/activities-standard/location
[6] http://aiddata.org/maps
[7] http://devtracker.dfid.gov.uk
[8] http://wbi.worldbank.org/boost
[9] https://openspending.org
[10] http://sempla.prefeitura.sp.gov.br/orc_homenew.php

geographical references on them, generating visualizations for the data, specially a map with different colored pins. Figure 1 describes CMN tool's architecture.

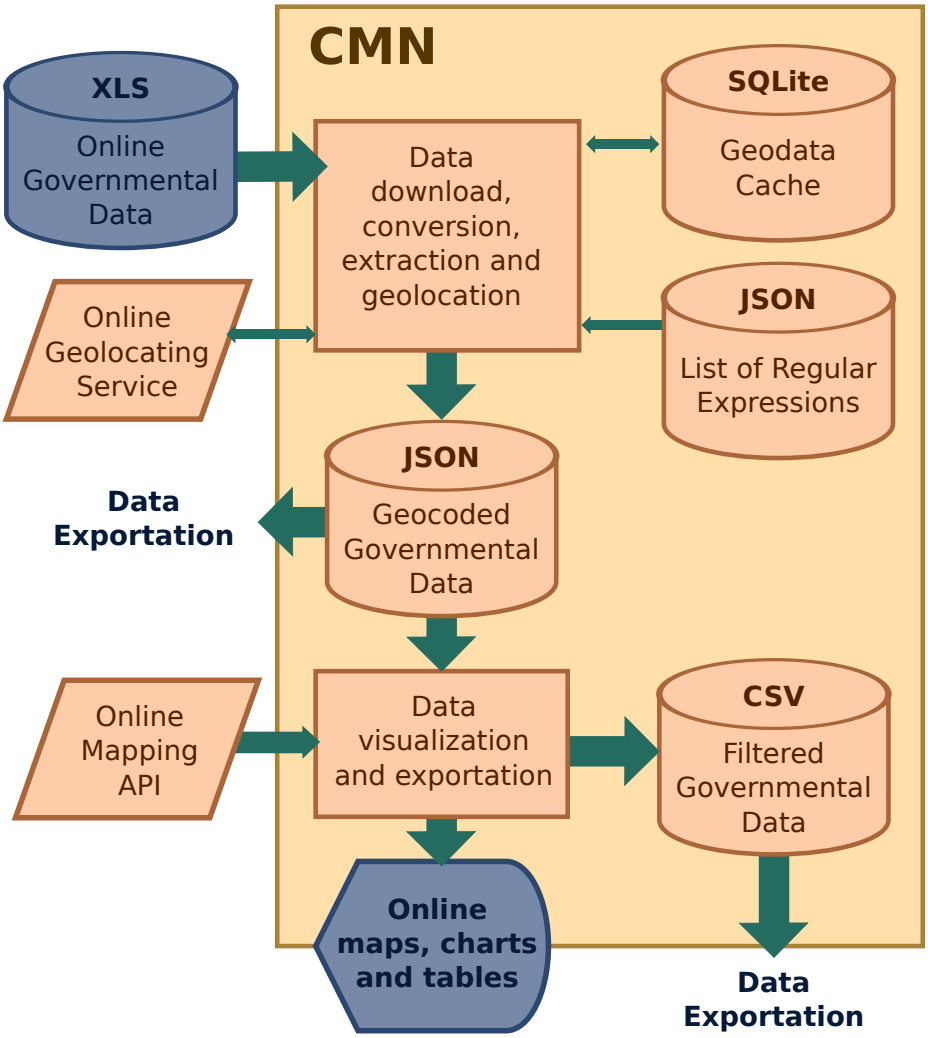

Fig. 1. Architecture of the CMN tool

The tool consists mainly of three scripts and a website. The first script, written in Python, downloads the XLS file from the government's website, converts the data to JSON and calls a second Perl script to process the converted data.

This second script applies a list of handmade regular expressions to the data, extracting elements that possibly describe a location, i.e.: "**street** São João" or "**school** Nova Esperança". The same script tries to geolocate these elements

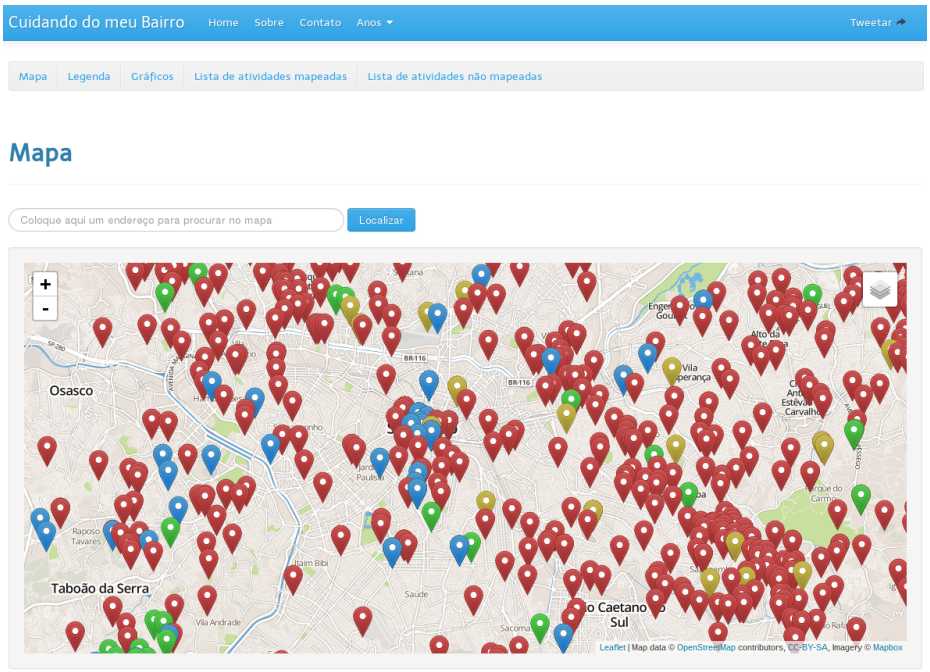

Fig. 2. Public expenditure map from the 2013 Budget in the City of São Paulo, generated by CMN

Fig. 3. Example of one public expenditure detailed information in CMN

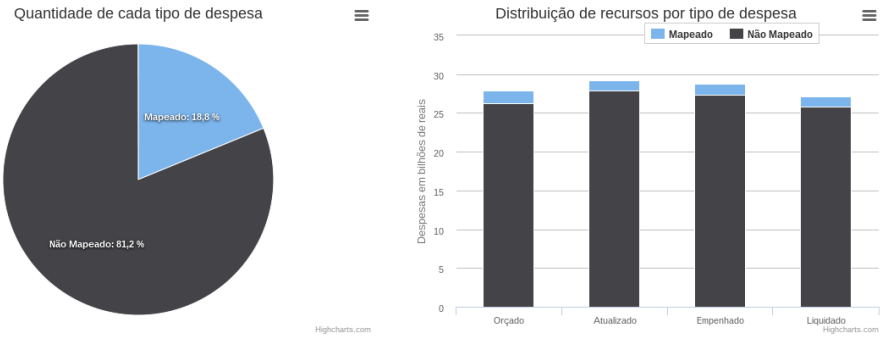

Fig. 4. Charts displaying the percentage of geocoded data

through an online service, saving the whole data (original governmental data and returned latitudes and longitudes) in a final JSON. A local cache, in SQLite, is used to avoid consulting more than once the latitudes and longitudes of an address. This way, if the initial XLS is updated by the government, when the scripts are run again to process the new data, only new addresses will need to be consulted online, speeding up the process.

The final JSON file and the third script, written in JavaScript, are served online via the PHP website. The script, when executed by a user's browser, loads the JSON file and process it generating three visualizations.

The first one is a map with a pin for each geocoded element. Figure 2 shows an example of a public expenditure map from the 2013 Budget in the city of São Paulo generated by the tool. Four colors were used to represent different stages of public expenditure, also exhibited in Figure 2. Budgeted (red color pin) means that it was directly included by a municipal agency or by city councilors amendments. Only after the expense is budgeted, it can receive the *commitment* (green color), which means the reservation of resources after the authorization or signature of service providing contract. Once the commitment is made and the service is performed, the expense is *liquidated* (blue pin). One extra color was included, (yellow color), which shows if there was any change in the budgeted value. It is important to observe that, since the legal framework enforces that all spending data must be updated in a 24-hour basis, CNM tool is configured to extract data from the government's website everyday and refresh the visualizations if needed.

As well as showing the expenditures in the map with their corresponding stages, the following information is presented to detail a specific expense, as shown in the example in Figure 3, and can be displayed by clicking on a pin:

- Identification number;
- Description;
- Budgeted, updated, committed and liquidated values;
- Agency responsible for that expense;
- Agency unit;

– Purpose of the expense, like education or health;
– More specific information on the purpose;

Finally, two charts present the percentage of geocoded data (Figure 4), in numbers of individual expenditures and in amount of money. It is important to display this comparison because there are many expenditures that cannot be placed, or some spending descriptions do not contain any information regarding a location where the money were spent.

The tool was able to identify spendings with a good level of description, however, they have symbolic resources attributed to them and did not progress until the committed or liquidated stages, only staying in the budgeted phase. On the other hand, there are others expenditures with large volumes of resources budgeted, committed and liquidated without any information to help the monitoring of that expense by the citizen. The authors list below the main problems and some examples of spending descriptions in São Paulo city budget:

– Aggregated information: there are projects and activities with large volume of resources that reach the commitment and liquidation phases. However, they do not provide enough details that allows expenditure public oversight. Example: "Operation and Maintenance of Public Libraries", without any detail about which libraries will receive the resource.
– Specific, but only stay in the budgeted stage: these expenditures have detailed description, but they do not progress to the committed stage. Example: "Building of the Municipal Kindergarten School in Vargem Grande neighborhood, in Parelheiros".
– Expenses with symbolic values: a very interesting situation where the tool placed a large number of pins about expenditures with small amount of money planned for them, despite their complexity to implement. These are projects or activities which have approximately 500 USD budget and this is not compatible to the service or good provided, e.g. "Building and Installation of the Municipal Hospital Parelheiros". Also these category of expenses do not reach the commitment phase.

In other to promote reuse of our data, the tool also allows the user to download the final JSON, which has the geocoded data. Besides that, it is also possible to export some filtered data in CSV to the users that want to analyse themselves the public spending divided by agency. In order to promote a wider replication and adoption of this tool, its source code has an open license and is available online[11].

4 CMN Project

This tool is being used by many social leaders and civil society organizations that comprises a network known as "Rede Nossa São Paulo" [12]. It gathers more

[11] https://github.com/okfn-brasil/cuidando
[12] http://www.nossasaopaulo.org.br

than 600 civil society organizations working in areas as diverse as education, health, housing, environment, security and leisure.

This project is also a case study in the context of the Open Data Research Network[13], supported by the World Wide Web Foundation and the International Development Research Center (IDRC) and its objective is to give valuable information on how access to budget information affects the relationship between civil society and public administrators in district and municipal levels.

Besides this, we are considering to mine other official sources of information in order to extract details about some expenditures. The main information source is the Official Gazette, which is a PDF document and contains legislation, jurisprudence and administrative actions. It contains a huge amount of information and describes, for instance, bidding and procurement which could substantively improve our project by providing elements for geographic location and other relevant information.

Nevertheless, a large amount of unstructured textual information is supposed to be processed before annotating and validating information that are publicly available. Some definitions found in the literature define semantic annotation as a specific schema to create and use metadata, enabling new methods of access to information [11]. Important related works are OpenCalais [12], Zemanta [4], Ontos [5], TextWise [6], LexML Project [7] and SIOP Project [8].

Another source of information comes from São Paulo City Council, which are the amendments to the city budget. These unstructured information may be parsed and mined in order to build an amendments map, possibly as a layer in CMN website. We also believe that some interesting analysis from 4-years datasets (the period of a city councilman's mandate) may show geographical and political influence in the city. This also may give interesting information about the legislative and executive powers dialogue in order to transform will (amendment) into action (committed and liquidated activities).

The forth and last part comprises support for social interaction. Many alternatives are being considered: interface to social networks, a crowd sourcing platform, and a mobile application. The definitions will rise from discussions, interviews and surveys that are being conducted in the context of our case study in the Open Data Research Network previously presented.

Figure 5 summarizes all the ideas and the ongoing work presented so far.

5 Conclusion

In this paper CMN tool was presented and discussed. With budget geocoding provided, projects are no longer just numbers in a report, but actual improvements or programs down the street from citizens' homes. Comparison between neighborhoods also becomes possible with this mapping tool. More generally, this work presents another way of presenting budgetary information instead of just digitizing existing reports and placing them online, contributing to make data

[13] http://opendataresearch.org

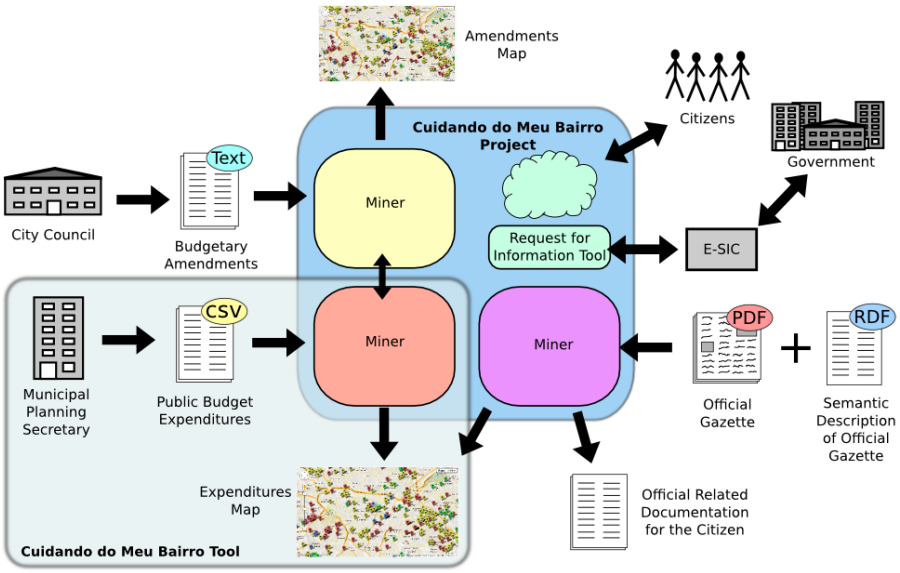

Fig. 5. An overview of CMN Project

more accessible and relevant to citizens. Basic visualizations, such as graphs, provide quick visual comparisons of datasets while maps provide geospatial references. Some related projects presented in this paper provide some or both graphical budgetary visualizations, but this work is the only that shows in a real time basis the city public spending. It makes possible to track the expenditure status, empowering the citizenship and enabling better budget oversight. We also understand that public officials do not have a similar tool and this also could be used to help public administration to be more efficient.

Also a broader scenario of CMN Project was briefly described. Different sources of information, most of them unstructured, are considered in our ongoing work: many of them come from official entities (executive and legislative powers of São Paulo city), but others may come from user interaction (comments, photos, requests for information, etc.). We aim to combine public sector information and crowdsourced data. We hope that this work can show how public information, particularly municipal budgetary data, may influence the relationship between civil society and the government. Moreover, the aim is to bring elements into debate that subside guidelines for building both to civil society and government in order to improve the control mechanisms and monitoring of public resources and help the fight against corruption.

References

1. Lei Complementar nº 101. Brasil (2000), http://www.planalto.gov.br/ccivil_

03/Leis/LCP/Lcp131.htm
2. Open Government Data Principles (2007), http://www.opengovdata.org/home/8principles
3. Lei Complementar nº 131. Brasil (2009), http://www.planalto.gov.br/ccivil_03/Leis/LCP/Lcp131.htm
4. Zemanta service (2009), http://developer.zemanta.com/media/files/docs/zemanta_api_companion.pdf
5. Ontos Semantic Technologies (2010), http://www.ontos.com
6. TextWise (2010), http://www.textwise.com
7. LEXML: Controlled Vocabularies (2011), http://projeto.lexml.gov.br/documentacao
8. Brandão, S.N., da Silva, T.S., Rodrigues, S., Araujo, L., Silva, D., de Souza, J.M.: Siop-legis: Thesaurus for selection and management of brazilian treasury domain. In: KMIS. pp. 195–200 (2011)
9. Craveiro, G.S.: Where does the money go? The challenge of transparency in state spending. Global Information Society Watch 2012 - The internet and corruption (2012)
10. Craveiro, G.S., Santana, M.T., Albuquerque, J.P.: Assessing Open Government Budgetary Data in Brazil. In: ICDS 2013, The Seventh International Conference on Digital Society. pp. 20–27 (2013)
11. Popov, B., Kiryakov, A., Kirilov, A., Manov, D., Ognyanoff, D., Goranov, M.: Kim–semantic annotation platform. In: The Semantic Web-ISWC 2003, pp. 834–849. Springer (2003)
12. R, T.: OpenCalais Documentation (2010), http://www.opencalais.com/documentation/opencalais-documentation
13. The 8 Principles of Open Government Data: (2007), http://opengovdata.org
14. Walker, S.T.: Budget mapping: Increasing citizen understanding of government via interactive design. 2014 47th Hawaii International Conference on System Sciences 0, 1–9 (2010)