

Experiences Integrating Heterogeneous Government Open Data Sources to Deliver Services and Promote Transparency in Brazil

Kellyton dos Santos Brito¹³, Marcos Antônio Silva Costa²

¹DEINFO - Pernambuco Rural Federal University

²Federal Ministry of Prosecution
Recife, Brazil

ksb@cin.ufpe.br, msilvacosta@mpf.mp.br

Vinicius Cardoso Garcia³, Silvio Romero de Lemos Meira³⁴

³Cin - Pernambuco Federal University

⁴Recife Center for Advanced Studies and Systems
Recife, Brazil

vcg@cin.ufpe.br, silvio@meira.com

Abstract—The use of Government Open Data to deliver services and promote transparency is a current goal of many governments. In this paper, we briefly describe two contest winner applications with focus on these goals, and present a discussion about the difficulties of integrating Government Open Data to develop them.

Keywords—open data, open government, e-government, transparency, digital society, Brazil

I. INTRODUCTION

Since the 50s, governments agree that to open government data to be verified and used by the general population, has the potential to increase accountability, citizen participation, and collaboration, while offering better public services, increasing efficiency and effectiveness, and decreasing corruption [1][2][3][4]. Currently, with the advent of new technologies, governments have made efforts to take advantage of the new "Open Government" scenario, and have launched initiatives such as the Open Government Partnership (OGP).

Brazil, one of the OGP founders, is very engaged to Open Data initiatives. The main consequence of its participation is the development and effectiveness of its Access to Information Law, which established a legal framework of guidelines for opening data from all levels of the government, including more than 5,500 municipalities. As a direct consequence, all levels of Brazilian government are launching Open Data Portals, and promoting application contests and hackathons to incentivize the population and non-governmental organizations to create applications that achieve Open Government objectives.

In this context, this paper briefly presents two applications: My National Congress (*Meu Congresso Nacional* - www.meucongressonacional.com) and Recife Citizen (*Cidadão Recife* - www.cidadaoecifense.com.br). Both are based on heterogeneous government open data sources and share the same architecture. In Section 2, we present application functionalities and architecture, and in Section 3 we

present the main lessons learned and future directions.

II. THE APPLICATIONS

Both applications started their development in the context of an applications contest. *Meu Congresso Nacional* was developed on the First Brazilian Parliament Hackathon, promoted by Chamber of Deputies in November, 2013. The event's main objective was to create a collaborative channel between the Brazilian Federal Parliament, the society, and the development of solutions with Parliament's open data. This application received first prize.

The second application, *Cidadão Recife*, was developed to participate in the Recife Application Contest called Smart Citizen (*Cidadão Inteligente*), promoted by Recife City Hall. Its objective was to incentivize developers "to build applications or conceptual projects that enhance life in Recife". This application also received first prize.

A. Applications functionalities

Both applications were developed to achieve the main benefits of Open Government Data: (i) to improve transparency and accountability; (ii) to improve and offer new public services; and (iii) to promote citizen participation and collaboration.

Meu Congresso Nacional focuses on (i) and (iii) by obtaining, organizing, and presenting in a friendly format the data about Brazilian Federal Deputies (513) and Senators (81). It presents (a) basic demographic data, such as their name, state, political party, and contact data; (b) data about parliamentary activity, such as proposed laws and amendments to the constitution (very common in Brazil); (c) detailed data about their public expenses (about \$15,000) monthly, on average; and (d) data from companies that receive the payments presented on (c), including their real location and facade. This application is a promising tool to be used in the next Brazil elections, to occur October, 2014.

Cidadão Recife covers all three areas: transparency, services, and participation. It shows detailed and grouped data

This work is partially supported by the Brazilian Federal Ministry of Prosecution; the INES Institute, funded by CNPq and FACEPE, grants 573964/2008-4, APQ-1037-1.03/08 and APQ-1044-1.03/10; the Brazilian Agency (CNPq processes number 475743/2007-5 and 140060/2008-1) and a CNPq scholarship.

about the City Hall budget, such as expenditures by function (education, healthcare, culture, etc), by type, by government agency, or by companies that receive the payments. In addition, the application provide health, educational and cultural unit finders, such as hospitals, daycare centers, schools, museums, and theaters among many others, based on user location and preferences. Furthermore, it implements an online vaccination card, which reminds users about their vaccination dates. A new functionality of collaboration, fully integrated with social networks, is under development to allow fast user-feedback about finances or directly to the health, educational, or cultural unit.

B. Applications Architecture

Both applications were developed based on a common architecture. Due to the length restrictions of this paper, we describe application architecture in a very high level, as shown in Figure 1.

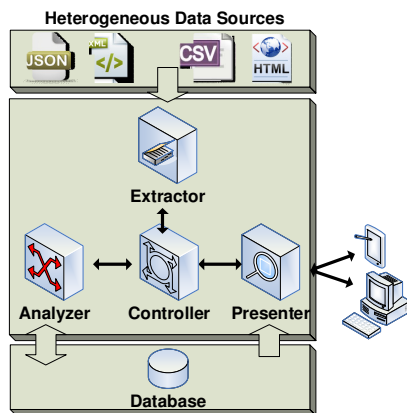


Fig. 1. Applications High Level Architecture

1) *Extractor*: Applications use several heterogeneous data sources. For each datasource, an extractor was developed with well-defined outputs, allowing them to be integrated on the application. The most important data sources are shown in Table 1.

It is important to note that some data sources are not open data. For instance, the Brazilian Internal Revenue Service (Receita Federal) website has a CAPTCHA that avoids automatic data gathering. Thus, with the first access to company data, the application shows the original CAPTCHA to the user, in order to access required data.

2) *Controller, Analyzer and Presenter*: These components are responsible to control application flow, analyze gathered raw data to provide useful information, and to show this information to the end user, respectively.

The applications are responsive and work correctly on desktop and mobile devices, and both have a plugin to automatic translation for nearly all of the popular languages.

III. LESSONS LEARNED AND FUTURE WORKS

By the development of both applications, some challenges can be identified:

- Multiple and decentralized data sources: Although a national open-data portal exists, there are many datasets that are not integrated on federal initiative. Thus, to build meaningful applications, a lot of datasets must be searched, analyzed, and used.

TABLE I. DATA SOURCES USED BY APPLICATIONS

Data	Type	Data Source
Deputies general data	json	Chamber of Deputies Open Data Portal
Deputies activity data	json	Chamber of Deputies Open Data Portal
Deputies expenditures	xml	Chamber of Deputies Open Data Portal
Senators general data	json	Federal Senate Open Data Portal
Senators activity data	json	Federal Senate Open Data Portal
Senators expenditures	csv	Federal Senate Open Data Portal
Companies general data	html	Internal Revenue Service website*
Companies location	xml	Google Geocode API
Recife city data	csv, json	Recife Open Data Portal

- Lack of standards for data publishing: In addition to the variety of data sources, each publisher chooses what and how to publish. There is no compliance with the national repository and no standard among cities, or even within the same portal or government agency. Thus, it is difficult to reuse software components that gather data from one source in another repository.

- "Zombie" data: In some cases, data is not "live" but only dumps of databases in a downloadable format. Even if accessed through an online API, there is no guarantee about source updates and none of portals explicit data update policy. The authors have verified that with best efforts, some datasets are updated daily. Thus, it is not possible to build real-time services.

- One-way data: Governments tend to publish one-way data from the government to the population. Applications can show such data, but the most useful applications, like online services where people return data to the government, are still limited.

Both applications are online and running, and are being improved for use in the next Brazilian elections and big events in Brazil, such as the 2014 Brazil World Cup.

REFERENCES

- [1] M. Francoli, "What Makes Governments 'Open'?", *JeDEM - eJournal eDemocracy Open Gov.*, vol. 3, no. 2, pp. 152–165, 2011.
- [2] J. C. Bertot, P. T. Jaeger, and J. M. Grimes, "Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies," *Gov. Inf. Q.*, vol. 27, no. 3, pp. 264–271, 2010.
- [3] D. Linders, "How can open development improve the effectiveness of aid?: leveraging open data, open standards, and web 2.0 interactivity for better development outcomes," *Proceedings of the 13th Annual International Conference on Digital Government Research*. ACM, College Park, Maryland, pp. 155–164, 2012.
- [4] S.-Y. Hung, C.-M. Chang, and T.-J. Yu, "Determinants of user acceptance of the e-Government services: The case of online tax filing and payment system," *Gov. Inf. Q.*, vol. 23, no. 1, pp. 97–122, 2006.