Use of Dublin Core to Increase Public Transparency of Brazilian Senate's Bills Datasets

Fernando de Assis Rodrigues  
Universidade Estadual Paulista, Brazil  
fernando@elleth.org

Ricardo César Gonçalves Sant'Ana  
Universidade Estadual Paulista, Brazil  
ricardosantana@marilia.unesp.br

Keywords: Open Government Data; data gathering; Metadata; Dataset; Brazilian Senate

1. Introduction

The transparency of government actions in society is an integral part of discussions about public administration models. These new public management sets seeks to redistribute skills and resources among different within and outside government organizations, allowing an increase in institutional pluralism on public office (Malin, 2006; Sant'Ana & Rodrigues, 2013). One way to strengthen transparency of government actions and to ensure a greater visibility of their activities can be achieved through an expansion of information-sharing environments that among its features provides new information flows between government and society (Rodrigues, Sant’Ana, & Ferneda, 2015). Thus, citizen participation will be extended beyond elections processes and the government will be able to improve their effectiveness and monitoring activities and results of their own actions (Bohman, 2000; Open Government Partnershi, 2014).

The Brazilian government establishes citizen rights to claim and access government information and data on a specific legislation called Information Access Law (from Brazilian Portuguese: Lei de Acesso à Informação). This legislation makes mandatory an use of Internet as a dissemination tool, towards to citizens grant access of Brazilian government's data (Brasil, 1988, 2011). Also, it is important that government datasets be machine-readable (Berners-Lee, Hendler, & Laslia, 2001) and available in a way that can be “[... suitable for use without re-typing or additional treatments to a direct data gathering [...]” (Sant’Ana & Rodrigues, 2013, p. 51) by external agents, independently of an initial format or a specific technology platform.

The goal of this paper is to presents an ongoing study of the applicability of using Dublin Core metadata terms in data retrieval to describe Senate's bills datasets, in order to increase the total amount of describing elements available to government data in gathering process.

The methodology adopted is based on an exploratory analysis of government datasets that were available on the set of Brazilian Senate's websites, on January, 2015. This analysis is divided into three phases: i) search for available bills datasets in Senate websites; ii) explicit metadata elements, already available in retrieved datasets; iii) find available information on these websites, specifically in datasets retrieval area pages, that can be part of a future Dublin Core metadata set of elements.

2. Website and dataset characteristics

The Brazilian Senate has a specific website to share data about its activities called “Portal e-Cidadania – Dados Abertos”. The website have forty five sets of data available, grouped into eight predefined groups. The information resource which contains the government datasets with bills and votes data is called Nominal Bills, and it is located in the 'Plenary Sessions' group. Nominal Bills consists on a set of twelve items, as follows: one description page; nine dump files for download in eXtensible Markup Language (XML) format, containing Nominal Bills data, grouped by year; one hyperlink that redirects users to other Senate's web site for queries, acting like a search interface to citizens; one hyperlink to a web service interface that provides an gateway to an external automated gathering process.
For each item, the website offers an unique page in HyperText Markup Language (HTML) format, divided by sections, as follows (in a top-down order): a) a paragraph with an item description; b) a 'download' button; c) a 'quick information board' with four elements (named: Part of dataset, Last updated, Format and License); d) an Additional Information (in a HTML table format) with two columns (named: Field and Value, not sortable) and nineteen rows with its field names written in lowercase and without blank spaces. All 'quick information board' elements and Additional Information field names are written in English language.

In web service interface for data gathering, it's possible to query Nominal Bills that occurred only by inputting a specific date. For example, to an external application collect these datasets and retrieve all Nominal Bills on a particular month, will be necessary to run 'x' queries, where 'x' represents a total of days in that month.

All queries results are in a XML format, and its elements hierarchy is organized as follows: a root attribute called BillsList (ListaVotacoes), who has two children elements: a) Metadata (Metadados) and b) Bills (Votacoes).

The Metadata element has a fixed number of children elements (three): a) Version (Versao) with its value being the date inputted previously as query parameter; b) ServiceVersion (VersaoServico), an integer with no further description, and; c) DataSetDescription (DescricaoDataSet) with its value for all queries being a fixed text (a sentence about the web service, with information about data updates and two hyperlinks (to a XML file and a XML Schema Definition (XSD) file). On Bills element, each children element represents data from a unique Bill result.

3. Results

When Nominal Bills datasets are gathering by external agents, it is available three metadata elements: Version, ServiceVersion and DataSetDescription. This set remains equal in web service interface and XML dump files. Other two hyperlinks were found in DataSetDescription element value – both redirects users to a Senate's web site error page.

All 'quick information board' elements and all rows in Additional Information section don't have any kind of explanation about its meaning.

4. Conclusion

In Additional Information sections, even rows doesn't have any kind of explanation about its meaning, it is possible that this set had potentially metadata information about Nominal Bills dataset that isn't available yet on retrieved files. For example, a value of 'name' field seems to be a description of dataset content; a value of 'id' field seems to be a unique resource identifier to dataset; etc. That kind of information could be more explored and added as a children element in Metadata element on XML files.

It concludes that on an application of Dublin Core set of descriptive elements on Senate's bills datasets have to observe the following variables: a) a study of meaning of elements found on unique pages; b) development a strategy that takes into account to fill Dublin Core required elements with values already available in Metadata's children elements and in dataset page section's, including an evaluation of adoption external software tools or data conversion algorithms on this process; c) a replacement of existing Metadata's children elements and namespaces.

As future work, it proposes a development of a dataset prototype with an application of the Dublin Core elements in a Nominal Bills XML dump file.
References