The Brazilian Federal Budget Ontology – A Semantic Web Case of Public Open Data

Luís Sérgio de Oliveira Araújo
Secretaria de Orçamento Federal
(Budget Federal Office) – SOF/MP
SEPN 516, bloco D, lote 8
55-61-20202232
luis.araujo@planejamento.gov.br

Mauro Tapajós Santos
Secretaria de Orçamento Federal
(Budget Federal Office) – SOF/MP
SEPN 516, bloco D, lote 8
55-61-20202119
mauro.santos@planejamento.gov.br

Daniel Aguiar Silva
Secretaria de Orçamento Federal
(Budget Federal Office) – SOF/MP
SEPN 516, bloco D, lote 8
55-61-20202058
daniel.aguiar@planejamento.gov.br

ABSTRACT
This paper presents and explains the workings and structure of the Brazilian Federal Budget Ontology, a web semantic model which aims at enabling developers and public finance experts full on line access to brazilian national (federal) budget data. The paper has an interdisciplinary flavor and is directed to both public finance and web semantic experts[1].

Categories and Subject Descriptors
I.2.4 [**Artificial Inteligence**]: Knowledge Representation Formalisms and Methods – *Semantic networks.*

General Terms
Design, Economics, Standardization.

Keywords
Pubic Budget, Model, Ontology, Semantic Web, Linked Data, Open Data, Transparency, OWL, RDF.

1. INTRODUCTION
This paper presents to the english speaking audience the Brazilian Federal Budget Expenditures Ontology. This model has been used by the brazilian federal government to publish its budget data in open data format. This ontology was first introduced in a portuguese language paper (Araujo et al. 2012), but since then it has been slightly improved so that data spanning from the years 2000 to 2015 could be represented. Also, this experience seems to be innovative and illustrative due to the potential usefulness of the model that can be adapted for utilization in other similar contexts, including by other countries. Also it is to be considered the ubiquitous interest and importance of the information which could be published through this procedure and the high standard of openness it has achieved – ranging to the 5 star level of Berner's Lee criteria for open data: (1) available in the Internet, (2) as structured data, (3) in non-proprietary format, (4) using URLs to identify things and (5) linked to other people's data). This is an interdisciplinary article since its subject matter is both related to fields of public budgeting and semantic web technologies, so that the presentation is couched in rather colloquial terms so that it can be grasped by both groups of adressees.

There is an ongoing question that frequently arises when someone talks about ontology in the computer science meaning of the term, which is "Why ontologies?". Considering the reasonable amount of investment that is required to understand the underlying concepts and managing all the connected technological apparatus necessary to create and use ontologies, this question, after all, is not that much out of context.

Of course there is one established tradition in the artificial intelligence and knowledge representation communities about the necessity of explicitly representing the parts of the world that are of interest for the solution of some problem so that it can make it possible to computationally reason, process, communicate and exchange information on that universe of discourse, with all the potential utilities that can be derived from these efforts. One famous example is the The Gene Ontology (GO) Project, which is a collaborative effort that provides structured vocabularies for annotating the molecular function, biological role, and cellular location of gene products. That said, in this paper we do not want to quarrel on these interesting methodological questions – as the specialists in the field themselves seem to have not yet come to a consensus on the issue[2] – but rather to show how was it possible to solve one specific important problem in the realm of public finance with the aid of an artifact built in Ontology Web Language – OWL.

In 2011 it was enacted in Brazil the Information Access Law (Lei do Acesso à Informação, Law 12.527/2011). Two trends have motivated this law. For one side, the demmand for transparency in public administration and in the other side, the increased disposal of modern information technology which have created concrete conditions for the widespread availability of digital information over the Internet. Specifically, this law demands that it is the duty of public organizations to promote the publication of

[1] The opinions laid down in this paper are exclusive responsibility of the authors and should not be taken as official stance of the institution they are affiliated to.

MEDES'15, 25-29 October 2015, Caraguatatuba/Sao Paulo, Brazil
© 2015 ACM. ISBN 978-1-4503-3480-8/15/10…$15.00
DOI: http://dx.doi.org/10.1145/2857218.2857232

[2] For one often cited account on this topic see Guarino et al., 2009.

public interest information created by or under their custody. Also, the law states that to achieve this goal, the public organizations must provide official sites in the Internet, and that, among the requirements of these sites, they should offer automated access by external systems in open format, structured and machine readable (art. 8º, § 2º).

2. THE PUBLIC BUDGET

One important piece of information created by public administration is the public budget. The budget contains information on all the predicted and realized amount of revenues and expenditures for the period of one year for one public entity. Generally public budgets are enforced by law for public entities in all levels of government. In Brazil, there are three federative levels of government, namely, federal, state and local (federal, estadual e municipal). This means that there must be one federal budget, one budget for each state government and one budget for each local government.

Public budgets exist as a consequence of one universally accepted public finance rule: no public expenditure can be made without legislative authorization[3]. Formally, the budget is a statement of predicted revenues and expenditures. It contains words and figures that propose expenditures for certain objects and purposes; and figures that are attached to each budget line item. Those who prepare the budget presume that future events will be shaped by what is written in the words and figures stated in the budget. Only through observation, however, it is possible to determine if the predictions postulated turn out to be correct. Budgets can also be seen as mechanism for making policy choices and as a plan of work for those with the responsibility to implement it, although, it has been observed that there may exist gaps between intentions and accomplishments. (Wildavsky & Caiden, 1996).

Although generally budgets comprise basically a declaration of revenues and expenditures, their presentations varies among different countries. Some effort toward a suggested standardization has been made by the IMF[4] for the classification of functions understood as socioeconomic objectives that general government units aim to achieve through expenditures. This common classification proposal is an important step towards the construction of a universally agreed upon model that could enable cross-sectional analysis among different countries and regions of the world, at least at a more aggregated level. One advantage of using OWL language is the possibility of easily converting objects under one classification method to another once there exists a straight correspondence between the meaning of the two groups of categories. Anyway the question if there is already somewhere another budget ontology that could be somehow reused would be better coached in terms of if there is another budget system that is widely adopted and used. And as we consider above, the answer is obviously no.

Anyhow, the structure of budgetary systems remains an issue often discussed[5] by reformers, the main motivations of these proposals being to find the holy grail of public finance, that is, the best way to make the budget, or the way to make the best budget. But as more recently it has been widely accepted[6], the question 'on what basis can it be decided that some amount of money should be spent on program X instead of program Y?' can not be answered on scientific grounds, being rather a matter in the realm of politics.

3. THE ONTOLOGY

The Brazilian Federal Budget Expenditure Ontology is a model of the expenditure side of the budget. This delimitation is a design decision aimed first at focusing in one demanding aspect of the budget for the purpose of strategy – divide and conquer, preserving the usefulness of the model. Future work may be devoted to the extension of the model to include the revenue side, just as to include other related important features such as, for example, information on the payments that were made backed with the authorization given by the budget.

The federal budget in its paper version comprises thousands of pages with tables and figures, so that it is almost unmanageable in its printed form. Some kind of electronic representation of the budget, e. g., in the form of a relational database is required so that it can be somehow manageable and therefore useful. The proposed model captures the essential aspects of the expenditure side of the budget, so that it turns out possible to selectively query for specific groups or kinds of expenditures. The model is general enough to allow for the representation of the brazilian federal budget data on multiple years.

The ontology was written in OWL language with the aid of the Protegè ontology editor. But the ontology alone, although a great accomplishment, would be of no avail without conforming data. So the Planning Ministry published the data on the federal budget expenditure for the years since 2000 to 2015 according to the proposed model we describe here. This data is available in the official open government data site, orcamento.dados.gov.br/sparql, in RDF format.

Basic to understanding the expenditure side of the budget is the classification system its structured upon. The expenditure classification system has manifold purposes, but basically there are four general goals: 1) program formulation; 2) contribute for budget execution; 3) assign responsibility; and 4) allow for the analyses of the economic effects of public activities (Burkhead 1956). Essentially, the classification system's purpose is to classify or group one amount of money according to various criteria. In Brazil the classification system has been evolving throughout decades and has reached relative stability, so that the last significant reform was adopted in 1998 (Decree 2.829).

We can view the expenditure side of the budget – and this constitutes the core of the model – as simply a list of items with values attached to them. Each line item (Item de Despesa) corresponds to one budget line in the paper report and is classified according to several criteria. These criteria are

[3] The U.S. Constitution (Article I, section 9, clause 7) states that "No money shall be drawn from the Treasury, but in Consequence of Appropriations made by Law; and a regular Statement and Account of Receipts and Expenditures of all public Money shall be published from time to time."

[4] The COFOG – Classification of the Functions of Government was produced by the Organisation for Economic Co-operation and Development (OECD) and published by the United Nations.

[5] Wildavsky (1996, Chapter 10, Reform, p. 264).

[6] That question, posed by V. O. Key, Jr. in his classical “The Lack of a Budgetary Theory” (1940), although several times attempted to be answered, none of the attempts met the challenge, as Wildavsky & Caiden (1996) say, for one excelent reason: “The task, as posed, is impossible to fulfill”.

embodied in the notion of Classifiers and the meaning of these classifiers or, in the ontology parlance, their semantics, is well established in the community of budget experts and users and can be found in LDO (Brasil, 2015a) and MTO (Brasil, 2015b).

Classification is a mental process in which we compare things and try to group them in homogeneous categories. The public budget process as an institutional collective endeavor is governed by established formal rules and is the result of this rigorously regulated process. Once we face the budget as it is put, that is, as observers, we want to execute the reverse mental process (from those who built it) and capture the information that is encoded in the budget by the classification system. In other words, given one line item, we want to know how the associated figure is classified – is this amount of money to be used for Education or for Defense? What public organization is accountable for this spending? Is it destined to be spent on current or capital spending? More broadly, we may also want to ask questions involving all budget line items – what is the total yearly spending destined to personel? How much is to be spent to serve public debt interest?

Given that we are talking about grouping things with similar characteristics – that is, a classification process, it is important to mull over what kind of stuff we are grouping or classifying together in chunks. Public budget information is composed of revenue and expenditures estimates of money to be collected and spent in a certain period of time, usually a year. Differently from private entities which use their own (or borrowed) resources to finance their everyday ongoing business, public entities rely on taxes, contributions and other public incomes and these financial resources can only be used if legislatively authorized, that is, enforced by law. This constitutional command, as it was said before, is one of the main functions of the budget. So while accounting in the private sector is mainly focused in the entity's assets and liabilities – and their variation –, accounting in the public sector is mainly focused on the budget process.

While a balance sheet enrolls assets and obligations, the budget enrolls revenues and expenditures, their predicted and executed values. The budget is financial in character, not patrimonial. The problem it faces is how to finance public entities activities in the short term. So liquidity is one important aspect of the financial resources that are to be available as revenues. Only liquid assets can be considered as revenues. Think about the assets side of a balance sheet in the private sector. The accounts are listed in increasing inverse order of liquidity. It is a big deal if the composition of the assets is mostly liquid (bank accounts, financial assets etc) or mostly illiquid (fixed assets). In public budgets, only liquid assets are considered as revenues. Non-financial assets can not be used to pay salaries or everyday expenditures in public agencies, unless they are previously converted to money through financial inversions. Also, since budgets are prepared in advance of their period of execution, their initial revenue value is of course an estimate and may not be realized as predicted and this poses one more degree of complexity to the process. Unpredicted contingencies may distort the initial prospects or, what is more critical, the very initial proposed prediction itself may not be universally agreed on.

Table 1. Public Budget Execution Information

Revenues		Expenditures	
Predicted	Realized	Predicted	Realized
1,000,000.00	950,000.00	1,000,000.00	970,000.00

Money is the primary stuff the budget is concerned with, since it represents liquid value that can be immediately converted in services and products destined to satisfy the necessities of the State – and therefore of the population. But different uses can be accomplished with money. So essentially expenditure budget classification is the classification of the use of the money: used by who, used for what purpose, used to create what output, used to buy what kind of input.

If spending can be classified, separated, or grouped it is because these groups have common characteristics that distinguishes them. So money spent is distinguished, e. g. according to which organization unit is responsible for that expenditure (institutional classification), according to what the purpose of this spending is (functional and programmatic classification), according to the kind of input that is used (spending element), according to the predicted economic impact of the spending (economic category) and according to other classifications criteria that exist to attend specific information needs. All this segregations occurs simultaneously. We can imagine a similar scenario for the sake of clarifying this multiple classification concept. Consider that we have a group of people, and that each person can be classified according various criteria like religion belief, education level, political philosophy, sex, nationality etc. Them the same person will be grouped in different categories. The situation is similar to what occurs with budget line items relative to the classification system.

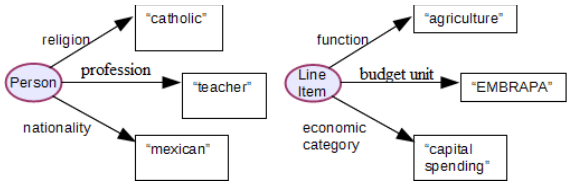

Figure 1. Multiple ways of classifying the same object

Now we can better grasp this interpretation process by considering one particular case of public spending similar to what occurs thousands of times during a fiscal year in public administration expediency. Lets pick one real spending case. In 2014, one agency in the Ministry of Agriculture decided to buy one truck vehicle. The value payed was R$ 356,828.50 reals (approximately equivalent US$ 105,000.00 dollars). This spending was predicted beforehand and estimates for the purchase were included in the budget. This means that some budget line in the budget had authorization value sufficient to cover the purchase. It is important to notice that one line item corresponds to several payments, all of which are identically classified since they are included in the same line item. The object of the classification process is the budget line item, not each particular payment. The following table summarizes the classification for the line item corresponding to this specific purchase.

Table 2. Acquisition: truck vehicle by Ministry of Agriculture

Classifier (English)	Description	Code
Sphere	Fiscal or Social Security	'01' – federal budget
Organization	Higher level of administrative agregation	'20' – Agriculture
Budget Unit	Lower level of administrative agregation	'22202' – EMBRAPA (Agriculture Research

		Brasilian Enterprise).
Function	Major purpose of the spending, such as Education or Defense.	'20' - Agriculture
Subfunction	Subdivision of the Function.	'572' –Technologic Development and Engeneering
Program	Group of structured actions aimed at solving some public problem.	'2042' - Inovations for Agricultures
Action	Actions are more specific and usually exist to provide some public good or service.	'20Y6' –Technology R&D for Agriculture
Subtitle	Indicates the location of that part of an action.	'0001' – National
Budget Plan	Another subdivision of action for diverse administrative purposes.	'0002'
Use identifier	Indicates if the value is to be used in conjunction with loans.	'0'
Revenue Source	Indicates the origin of the revenue that corresponds to this expenditure.	'100' – Ordinary resources
Economic Category	Can be eather current or capital.	'4' – Capital expenditure
Expenditure Group	Specifies one of 7 categories of expenditures.	'4' - Investiment
Aplication Type	Indicates if the spending is to be transferred to other entities.	'90' – Direct applications
Spending Element	Indicates the type of input which is been purchased.	'52' – Equipamento e material permanente
Primary Result Indicator	Does it affects the primary result (non-financial revenue and spending)?	'2'

Since the budget expenditure side can be seen as a list of line items, the ontology simply reflects this aspect. There is one basic class called ItemDespesa (Line Item). The members (or instances) of this class are the budget line items. The other classes in the model represent the classifiers. So there are 15 classifiers. All classifiers are defined as subclasses of the class Classifier (Classificador). This strategy was pursued so that the two properties of Classifier class – code (codigo) and description (descricao) could be reused by all classifiers. Another class introduced was the year (exercicio) class which represents the fiscal year of the concerned budget. The following figure is a simplified representation of the ontology. For a complete version of this diagram see http://vocab.e.gov.br/2013/09/loa.

It is important to remember that choosing to model the budget with an OWL ontology entails the opportunity to expand it later to include more sophisticated relations between the concepts, something you could not do with a simple taxonomy.

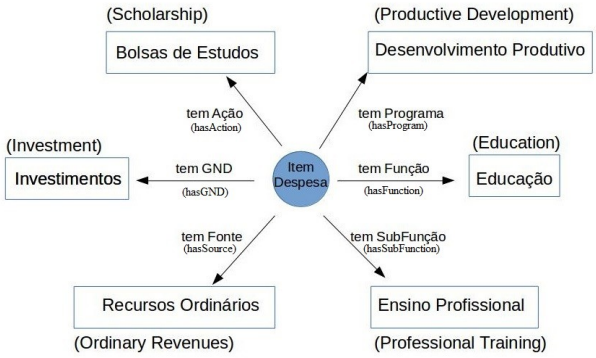

Figure 2. Ontology

4. RDF - RESOURCE DESCRIPTION FRAMEWORK

This budget ontology can be represented in RDF language, which is the W3C standard format for data interchange on the Web. RDF extends the linking structure of the Web and uses URIs to name the things and the relationship between things as well. Resource is the term used in the Internet parlance to designate anything that can be represented in the web. A person, an organization, an event, a web page, a product, a book, an abstract idea. Every resource can be designated by one unique URI. Here is a couple of example URIs.

http://orcamento.dados.gov.br/doc/2014/ItemDespesa#N65631

ftp://example.org/resource.txt

The RDF model has the form of a simple declaration that contains three elements: the subject, the predicate and the object. And this structure corresponds to our everyday communication parlance ("John plays tennis"). In RDF each element is represented by one unique URI, including the predicate. This turns it liable to be published in the Internet. RDF databases are composed of thousands and sometimes billions of triples. There are available free and proprietary tools designed to store, convert relational databases and publish data according to the RDF paradigm. RDF also provides more advanced concepts that can be used to structure the information, like the concept of classes so that it can be possible to establish constrains on the declared data. One can say that 'John' belongs the class 'Person' and that only instances of the class 'Person' plays tennis. This is sufficient for a declaration like 'USA' plays 'tennis' to be rejected on the grounds of been inconsistent with the constraint imposed by the respective knowledge base (ontology). This process of classification can be quite sophisticated and in fact there is one burgeoning field research in computer science that focuses in the abstract characterization of the objects and classes in any possible context, which is the study of ontology as technological artifacts. This field of research has a strong inspiration in the tradition philosophical discipline of Ontology and in fact has adapted many of its insights.

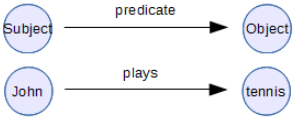

Figure 3. Triple

This declaration is usually referred to as a "triple". Using this simple model, RDF allows structured and semi-structured data to be mixed, exposed, and shared across different applications. The first structured step of RDF working process is to define classes of objects and that certain objects belong to certain classes. RDF enables the creation of large declarative knowledge bases that can be accessed publicly in the Internet, either directly by querying the databases or indirectly through the use of semantic web applications that can interact with these repositories.

RDF is part of a group of technological iniciatives and achievements known as Semantic Web which encompasses many other models, languages, frameworks and tools. Among the most useful in our context here is SPARQL query language, which is used to query triple storage repositories. SPARQL is fairly technical but very intuitive specially for those already experienced with database SQL language. To have a feel of how this works in practice, we show here a couple of examples of queries that can be made right on line in the federal brazilian budget. One can access orcamento.dados.gov.br/sparql and run the SPARQL query he wishies.

```
1 SELECT  (sum (?DotacaoInicial)) as ?soma   WHERE
2 {
3  ?linha  rdf:type  loa:ItemDespesa .
4   ?linha  loa:temExercicio  [loa:identificador 2015] .
5   ?linha  loa:valorDotacaoInicial  ?DotacaoInicial .
6 }
```

Figure 4. Sparql example

You can just copy and paste this query in the site www.orcamento.dados.gov.br/sparql and them click in the 'Run Query button to see the results. Now what it does is

Line 3: to select from the triple storage all the triples that are of the type loa:ItemDespesa (that represent budget line items);

Line 4: to select only the triples that have the value '2015' as the property 'temExercicio' (hasYear).

Line 5: to sum the values of the property valorDotacaoInical (initial value) of all triples and place the result in the variable ? soma (sum).

You can use this same query and change only the year, for example, from 2015 to 2014, and see the result for the year 2014.

Many other combination can be drawn. The basic knowledge which is necessary is the ontology of the brazilian federal budget, which is available at http://vocab.e.gov.br/2013/09/loa and the ability to use SPARQL query language, which fairly known in the semantic web community.

5. CONCLUSION

> "With the aid of modern devices which, if I am not mistaken, the practical Americans have already put in use, perhaps two or three seconds would be enough for this procedure."
>
> Knut Wicksell, 1896 – on the use of electrical devices in voting procedures.

This remark was made by the swedish economist Knut Wicksell in his Finanztheortische Untersuchungen (1896) – A New Principle of Just Taxation –, a work focused on the organization and functioning of the state. In this work Wicksell relates the benefit principle of taxation with the parliamentary approval of taxes, that is, the expenditure against the revenue side of the public budget. He clearly seems to express promising enthusiasm on the possibilities that technology can entail in the future political arena.

Wicksell is more famous for his contribution to economic theory dealing with markets. But his work on the functioning of the state is the more illuminating mainly because he refuted the mainstream economic orthodoxy of his time that considered the market as the only object of inquire of economic theory, the functioning of the state being kept aside as a kind of black box. His insights were later popularized by James Buchanan and others who gave rise to an active intellectual movement known as Public Choice School of Political Economy.

To the purposes in this work what is important to highlight is the role played by technology as a tool for political participation. We show how a set of technological artifacts can make it possible for ordinary people to have access to crucial public interest information produced by government as timely and as thoroughly as the public officials themselves. As one provocative political thinker points out, the striking feature of direct democracy is that, for the first time in human history, it is now technologically possible (Budge, 1996). But, it is recognized, this is not all that has to be done. Technology is a tool. To fully profit from its use it is necessary to acquire specific skills and knowledge. The open data that is put to the public by this ontology needs further user interface that can make it really accessible by the lay man. Initially, it was thought that this role could be voluntarily performed by social entrepreneurs (specialized social organizations devoted to public sector analysis)[7]. The observed results so far have shown that it may be necessary more work on the part of the government to explain and educate the public on using this artful tool.

6. REFERENCES

[1] ARAÚJO, Luís; SILVA, Daniel; SANTOS, Mauro; CRUZ, Fernando; FONSECA, Matheus; BERNARDES, Guilherme. Uma ontologia das classificações da despesa do orçamento federal. ONTOBRAS – MOST – Joint Seminar on Ontology Research in Brazil and VII International Workshop on Metamodels, Ontologies and Semantic Technologies. Recife, Brazil, 2012.

[2] Brasil (2015a) Lei 13.080, de 2 de janeiro de 2015. http://www.planalto.gov.br/CCIVIL_03/_Ato2015-2018/2015/Lei/L13080.htm

[3] Brasil (2015b) Manual Técnico Orçamentário – MTO. Edição 2016. Brasília, 2015. http://www.orcamentofederal.gov.br/informacoes-orcamentarias/manual-tecnico/mto_2016_1aedicao-200515.pdf

[4] BUDGE, Ian. The New Challenge of direct Democracy. Polity Press. Cambridge, UK, 1996.

[5] GUARINO, N; OBERLE, D; STAAB, S. What is an Ontology? Handbook on Ontologies 2009, 1-17.

[6] WICKSELL, Knut. [1896] A new principle of just taxation. London: New York: Macmillan, 1958. p.72 - 118 (Classics in the theory of public finance).

[7] WILDAVSKY, Aaron; CAIDEN, Naomi. The new politics of the budgetary process. 3.ed. New York: Longman, 1996.

[7] For one interesting initiative made by a high school student, see Web 21 Sep 2015. <http://gastospublicos.com.br>.