

Requisitos para construção de um padrão publicação da execução de orçamento público no Brasil

Marcelo T. Santana¹, Gisele S. Craveiro¹

¹Escola de Artes, Ciências e Humanidades – Universidade de São Paulo (USP)
Av. Arlindo Bétio, 1000 – 03828-000 – São Paulo – SP – Brazil

{marcelo.tavares,giselesc}@usp.br

Resumo. *O Brasil tem um quadro legal que impõe a todas entidades governamentais a publicação de dados detalhados da execução do orçamento em tempo real na Web. Conforme um diagnóstico encontrado na literatura, a grande maioria dos portais de transparência não atende o nível mínimo de qualidade da informação prevista no arcabouço legal, o que leva a uma situação de opacidade das contas públicas. Este artigo apresenta requisitos a ser considerados numa proposta de padronização de publicação de dados da execução orçamentária no Brasil e aponta alguns caminhos.*

1. Introdução

O acesso à informação das finanças públicas é essencial para a transparência nas ações do governo, para aumentar sua confiança e responsabilidade. Os movimentos da Nova Gestão Pública e da Nova Gestão das Finanças Públicas (Frederickson, 1996)(Pollitt, 2000) mostram a importância da transparência como um pré-requisito para a prestação de contas. De acordo com a OECD, 2002, "o acesso à informação, consulta e participação ativa na formulação de políticas contribui para a boa governança, fomentando uma maior transparência na formulação de políticas; mais responsabilidade".

O Brasil tem um bom e reconhecido quadro legal sobre transparência fiscal, como a Lei Complementar n° 131 (Brasil, 2009) que requer de entidades públicas brasileiras a publicação, em tempo real, de informações detalhadas sobre a execução orçamentária na *Web*. Na Lei Complementar n° 12.527 também se reforça que muitas informações do setor público, despesas, transferências financeiras e aquisições devem ser publicadas da forma primária, autêntica e atualizada (Brasil, 2011).

Em 2007, a Open Government Working Group publicou oito princípios de Dados Governamentais Abertos visando a preservação do direito de informação para toda sociedade. Neles os dados devem ser completos, primários, atuais, acessíveis, processáveis por máquina, não-discriminatórios, estar em formatos não-proprietários e terem licença livre. Um estudo recente examinou dados de execução do orçamento público em 88 portais de transparência no Brasil em relação aos oito princípios de DGA(Craveiro *et al*, 2013) e, embora em alguns aspectos tenha sido encontrada conformidade com esses princípios, uma das principais conclusões leva à falta de aderência ao princípio de primariedade dos dados, sendo apenas 17.16% de todos os conjuntos de dados analisados. Além disso, outro resultado mostra 30% dos conjuntos

de dados eram processáveis por máquina, mas é importante notar que isso é relativamente fraco, uma vez que na maioria dos casos os dados foram publicados de forma semi-estruturada, sem metadados nem vocabulário comum, pois os dados considerados processáveis por máquina nesse estudo foram encontrados no formato CSV ou XLS.

A Figura 1 resume a conformidade da amostra nesse estudo com os princípios de Dados Governamentais Abertos (DGA). Os autores observaram que apenas os princípios completo, acessível e não-discriminatório estavam em conformidade superior a 50%. Por outro lado, apenas 4,4% dos sites analisados forneciam informações atuais e nenhum apresentou licença livre para uso dos dados.

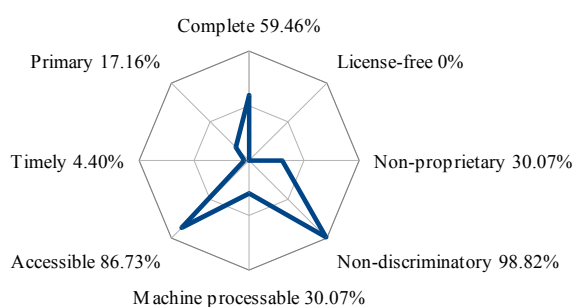


Figura 1. Conformidade da amostra com os princípios de DGA (Craveiro et al, 2013)

Estamos de acordo com Wang e Strong (1996) que a má qualidade de dados pode ter impactos sociais e econômicos substanciais e exigem que a conceituação dessa qualidade deva emergir a partir da perspectiva dos consumidores de dados. Este artigo visa dar uma contribuição nesta direção, apresentando alguns requisitos para publicação de dados de execução de orçamento público relativos ao caso específico da realidade brasileira.

2. Objetivos da Pesquisa

A pesquisa objetiva, a partir de um levantamento de requisitos da literatura sobre dados abertos, discutir as características necessárias para uma padronização de publicação de dados abertos de execução de orçamento público.

3. Contribuições Esperadas

O Brasil é uma federação com o tamanho continental e milhares de entidades, incluindo poderes executivo, legislativo e judiciário nas esferas federal, estadual e municipal. De todos os dados que são gerados por estas entidades, a Lei Complementar nº 131 (Brasil, 2009) só exige a publicação de dados sobre a execução do orçamento.

Tendo em vista a falta de aderência de dados do orçamento publicados com a abordagem de DGA espera-se contribuir para uma construção de padrão para publicação de dados que permita novas formas de reuso por máquina.

4. Resultados já alcançados

Da revisão da literatura de DGA e LD foram levantados requisitos a serem considerados na implementação de sistema de publicação de dados. Esses requisitos são apresentados nessa sessão e associados ao modelo de sistema de informação da Figura 2. Esse modelo em quatro camadas segue a abordagem de desenvolvimento *top-down* a partir do cliente, ou seja, a partir do usuário de dados do orçamento público, e neste trabalho é utilizado para auxiliar na identificação de requisitos para publicação de dados, ou seja, aqueles que devem ser implementados na camada de apresentação.

Bernes-Lee em 2009, desenvolveu um modelo de dados abertos ligados em cinco níveis (chamados de estrelas) para classificar progressivamente a facilidade de uso por consumidores de dados. Este modelo é uma classificação de referência de dados abertos na abordagem Linked Data (LD) e complementa os oito princípios de DGA. As três primeiras estrelas referem-se a licença-livre (*uma estrela*), formatos processáveis por máquina (*duas estrelas*) e formato não-proprietário (*três estrelas*), mas três questões são adicionados por este modelo:

- Estar disponível na Web (parte da classificação *uma estrela*): o princípio difere do princípio *acessível*, porque define que a via de acesso é a Web.
- Processável por máquina significa ser um formato de dados estruturado (*duas estrelas*);
- Identificar as coisas (*quatro estrelas*): "Usar padrões abertos da W3C para identificar as coisas, de modo que as pessoas podem apontar para as suas coisas." Criar um identificador único é um meio de permitir que outros conjuntos de dados podem ser ligados ao mesmo recurso.
- Ligar dados (*cinco estrelas*): "Vincular seus dados com os dados de outras pessoas para fornecer contexto."

Notamos que a publicação de dados no Brasil em relação ao modelo de Linked Data não pode ser classificado como uma estrela, porque, apesar de os conjuntos de dados de execução orçamental serem publicados na Web, eles não tinham licença livre declarada (Craveiro *et al*, 2013), apesar da Lei de Acesso a Informação (Brasil, 2011) garantir que todas as informações do governo são de domínio público *a priori*, o uso delas podem ser limitado juridicamente. Se considerarmos que dados estruturados são aqueles que possuem um esquema declarado, ou seja, informação de estrutura de dados, os conjuntos de dados não são estruturados, porque 49,27% dos conjuntos de dados de orçamento no Brasil foram publicados em CSV ou XLS, o resto em PDF (Craveiro *et al*, 2013), sem seguir um padrão de publicação. Apenas 30,07% dos conjuntos de dados estavam em CSV (Craveiro *et al*, 2013) e não podem ser considerados cinco estrelas pois esse formato não suporta *links*, assim, não é possível nesse cenário ligar dados orçamentários no Brasil.

Também em 2009, Hartig desenvolveu um modelo de proveniência de dados Web, que inclui duas dimensões - criação de dados e acesso a dados - a fim de "apoiar a avaliação das qualidades de dados, tais como precisão, confiabilidade e pontualidade". Por isso, acrescentamos a publicação de informações de proveniência como requisito.

Sheridan e Tennison, 2010, relatam a experiência do Reino Unido em publicar

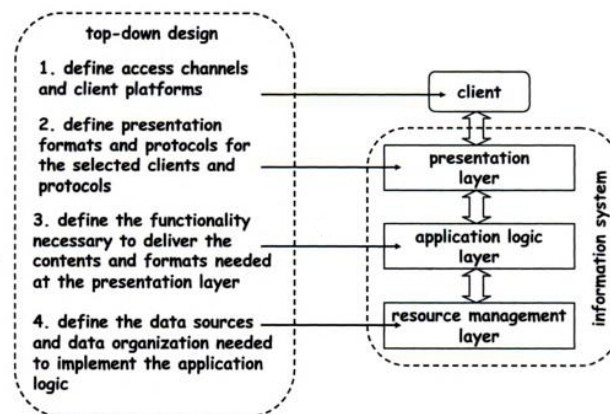


Figure 2: Information System model (Alonso, 2004)

dados como Linked Data e afirma que os formatos de dados como RDF¹ são "impenetrável sem analisadores especiais, tornando-se difícil usar [os dados] por não-especialistas". Devido a essa dificuldade, e para a ampla participação cívica, o governo do Reino Unido decidiu fornecer dados em formatos CSV² e XML³. Por isso, incluir o fornecimento de alternativas de acesso para a sociedade é também um requisito.

Open Knowledge Foundation, 2011, indicou a necessidade de se "organizar um catálogo central para listar as suas bases de dados abertos" com a finalidade de aprimorar a descoberta dos dados. Além disso, vocabulários controlados são uma forma de ajudar os cidadãos na busca de informações, pois "alterações na semântica de termos em vocabulários dados ligados podem ter uma influência dramática sobre a interpretação de conjuntos de dados" (Käfer, 2012). Zhu *et al*, 2012 afirma que a ligar dados não é suficiente, metadados são obrigatórios para um uso efetivo dos dados, pois proporcionam a criação de novas hipóteses ao melhorar a capacidade de se encontrar padrões nos dados. Metadados também podem ser utilizados para se formar coleções de dados (Powell, 2009).

A Tabela 1 lista os requisitos que foram levantados da literatura neste trabalho e onde deveriam ser implementadas no modelo de sistema de informação na Figura 2, onde cada requisito foi confrontado com as camadas considerando a origem dos dados e a experiência do consumidor de dados. Por exemplo, para os dados serem completos na apresentação é necessário serem completos no gerenciamento de recurso, caso contrário não é possível ao consumidor baixar esses dados. A licença livre não precisa existir na lógica de aplicação nem no gerenciamento do recurso, mas precisa ser apresentada ao usuário. O aspecto não-discriminatório depende exclusivamente da lógica de aplicação, que não deve selecionar por usuário quais dados no gerenciamento do recurso são apresentados.

Catorze desses requisitos devem ser representados na camada de apresentação. Sete requisitos referem-se a dados que devem existir na camada de gerenciamento de

1 www.w3.org/RDF

2 en.wikipedia.org/wiki/Comma-separated_values

3 www.w3.org/XML

Tabela 1 – requisitos levantados na literatura

Requisito	camadas do sistema de informação
Dados devem ser completos	gerenciamento de recurso, apresentação
Dados devem ser primários	gerenciamento de recurso, apresentação
Dados devem ser atuais	lógica de aplicação, apresentação
Dados devem ser acessíveis	lógica de aplicação, apresentação
Dados devem ser processáveis por máquina	gerenciamento de recurso, apresentação
Dados devem ser não-discriminatórios	lógica de aplicação
Dados devem estar em formatos não proprietários	apresentação
Dados devem ter licença livre	apresentação
Disponíveis na Web	apresentação
Identificador único dados	apresentação
Ligar dados a outros dados	lógica de aplicação, apresentação
Ter informações de proveniência	gerenciamento de recurso, apresentação
Dados devem ser permanentes	gerenciamento de recurso, apresentação
Prover alternativas de acesso	lógica de aplicação, apresentação
Usar vocabulários controlados	gerenciamento de recurso, apresentação
Usar metadados	gerenciamento de recurso, apresentação

Fonte: elaborado pelos autores

recurso, sem os quais não poderiam ser representados na apresentação. Alguns requisitos devem ser contemplados na lógica de aplicação, onde podem surgir novos dados, por exemplo, a informação de data e hora da criação do conjunto de dados na camada de apresentação; essa informação não existe no recurso mas deve aparecer como parte dos dados de proveniência.

5. Conclusão

Os requisitos apresentados neste trabalho são em si desafios que devem ser alcançados num contexto de publicação que não atende as abordagens de Dados Governamentais Abertos num Linked Data. O mapeamento desses requisitos com um modelo de sistema de informação auxilia na tomada de decisões de desenvolvimento de um ambiente de publicação de dados do orçamento público.

Com base nesses requisitos pode-se selecionar tecnologias para publicação de dados que garanta o direito à informação, através do atendimento aos princípios de DGA, e permita novas formas de reuso de dados por máquina pela inclusão de metadados, vocabulários controlados, informações de proveniência, e outros requisitos encontrados na literatura.

6. Referências

- Alonso, G. *et al.*: “Web services”. Springer Berlin Heidelberg, 2004.
- Berners-Lee, T.: Linked Data. 2009. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>.
- Brasil: “Lei Complementar nº 131”. 2009. Disponível em: <http://www.planalto.gov.br/ccivil_03/Leis/LCP/Lcp131.htm>
- Brasil: “Lei nº 12.527”. 2011. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm>.

- Craveiro, G. S., Santana, M. T. e Albuquerque, J. P.: "Assessing Open Government Budgetary Data in Brazil". In The Seventh International Conference on Digital Society 2013.
- Frederickson, H. G.: "Comparing the Reinventing Government Movement with the New Public Administration," *Public Administration Review*, 1996, 56(3), 263-270.
- Hartig, O.: "Provenance Information in the Web of Data". In *Linked Data on the Web. 2009*. Disponível em: <http://events.linkedata.org/ldow2009/papers/ldow2009_paper18.pdf>
- Käfer, T. *et al.*: "Towards a Dynamic Linked Data Observatory". In: *5th Linked Data on the Web Workshop at WWW. 2012*.
- OECD: "OECD Best Practices for Budget Transparency," OECD publications, France, 2002. Disponível em: <www.oecd.org/gov/budgetingandpublicexpenditures/Best%20Practices%20Budget%20Transparency%20-%20complete%20with%20cover%20page.pdf>
- Open Government Working Group: *8 Principles of Open Government. 2007*. Available: <<http://www.opengovdata.org/home/8principles>>
- Open Knowledge Foundation: "Beyond Access: The Right to (Re)Use Public Information" (p. 8-9). 2011.
- Pollitt, C.: "Is the emperor in his underwear? An analysis of the impact of public management reform," *Public Management*, 2000, 2(2), 181-199.
- Powell, L. F.: "Building a strong foundation with Dublin Core". In: *International Conference on Dublin Core and Metadata Applications. 2009*. p. pp. 87-95.
- Sheridan, J. e Tension, J.: "Linking UK Government Data. In *Linked Data on the Web*", 2010, Raleigh, North California.
- Wang, R. Y., Strong, D. M.: "Beyond accuracy what data quality means to data consumers", 1996, *Journal of Management Information Systems*, v. 12, n.4, p. 5-34.
- Zhu, H. *et al.*: "Data and Information Quality Research: Its Evolution and Future", 2012, In *MIT Information Quality Program Proceedings*. Disponível em: <http://mitiq.mit.edu/Documents/Publications/Papers/2012/Madnick_2012_Data%20and%20Information%20Quality.pdf>