

O papel dos Vocabulários Controlados em projetos de Governo Eletrônico – uma visão da situação brasileira

Jaime A. Pinto, Maurício B. Almeida

Programa de Pós-Graduação em Gestão & Organização do Conhecimento
Universidade Federal de Minas Gerais (UFMG)
Av. Antônio Carlos 6627 – 31270-901 – Belo Horizonte – MG – Brasil

{jaimepinto,mba}@eci.ufmg.br

Abstract. *Our subject of investigation here is the Controlled Vocabularies for Electronic Government — VCGE. First, we present the current Brazilian situation regarding controlled vocabularies, revisiting the regulations and strategic plans of the Brazilian government. Then, we address the issue of interoperability between government agencies and services, it to another issue, namely, the data heterogeneity. Finally, we present an analysis of VCGEs and a possible alternative for improvements through the reuse of a public domain vocabulary. This research is part of a thesis scheduled to be concluded in July 2020. In such a thesis, we conducted an applied research with exploratory objectives. The overall goal is to build an ontology of integration between business vocabularies and governments institutions.*

Resumo. *Nosso assunto de investigação aqui é o Vocabulário Controlado para Governo Eletrônico - VCGE. Primeiramente, apresentamos a atual situação brasileira em relação aos vocabulários controlados, revisitando os regulamentos e planos estratégicos do governo brasileiro. Em seguida, abordamos a questão da interoperabilidade entre agências e serviços do governo relacionando-a com o problema da heterogeneidade de dados. Finalmente, apresentamos uma análise do VCGE e uma possível alternativa de evolução através da reutilização de um vocabulário de domínio público. Esta pesquisa faz parte de uma tese a ser concluída em julho de 2020. Nessa tese, realizamos uma pesquisa aplicada com objetivos exploratórios. O objetivo geral é construir uma ontologia de integração entre vocabulários de negócios e instituições governamentais.*

1. Introdução

Este trabalho destina-se à apresentação no Workshop de Teses e Dissertações em Ontologias que será realizado em conjunto com o ONTOBRÁS 2018 – Seminário de Pesquisas em Ontologias do Brasil – no período 01-03 de outubro de 2018, em São Paulo. Apresenta-se a pesquisa em andamento no Programa de Pós-graduação em Gestão & Organização do Conhecimento – PPGOC – da Escola de Ciência da Informação – ECI – da UFMG. O professor orientador é o Dr. Maurício B. Almeida, PhD.

Tendo como base o cenário atual de planos, recomendações e necessidades do Governo Brasileiro em projetos de Governo Eletrônico – eGov – buscamos neste trabalho relacionar as contribuições que a Ciência da Informação pode oferecer neste momento. Nota-se, nos documentos governamentais, as referências feitas a construções de organização do conhecimento do tipo Ontologias, Taxonomias e Vocabulários

Controlados. Tomamos então esta frente de pesquisa, analisando as definições já existentes e suas possíveis evoluções. Como passo inicial, analisamos os Vocabulários Controlados.

O objetivo geral desta pesquisa é construir uma ontologia de integração entre vocabulários de negócio e governos. E, de maneira complementar, apresentar formas de utilização desta ontologia em projetos de eGov que envolvam integração semântica de dados heterogêneos e arquitetura corporativa.

Parece-nos bastante promissor a possibilidade de reuso de um vocabulário de domínio público, pelas seguintes razões:

- Há um número crescente de vocabulários em código aberto.
- O conhecimento aberto, desenvolvido por uma comunidade de usuários, tem chances muito maiores de atender aos critérios de utilidade e qualidade do que um desenvolvimento isolado.
- Os artefatos da Modelagem Semântica são, de modo geral, interoperáveis, permitindo acomodar as especificidades brasileiras e individuais de cada órgão ou organização usuária.
- Parece ser bastante adequado que o ePing não estabeleça normas próprias e sim enderece outros padrões e organizações internacionalmente reconhecidas.

2. Marco Teórico

Adotamos neste trabalho o conceito de Governo Eletrônico que refere-se ao uso pelos governos das Tecnologias de Informação e Comunicação – TIC – e, particularmente a Internet, como uma ferramenta para criar valor de interesse público¹ (UN, 2016).²

Uma característica do eGov que nos interessa particularmente é a sua heterogeneidade e necessidade de interoperação. O relatório da Organisation for Economic Co-operation and Development – OECD – (OECD, 2014) diz que: “... *(Governo eletrônico) depende de um ecossistema de governo digital composto por atores governamentais, organizações não-governamentais, empresas, associações de cidadãos e indivíduos que apoiam a produção e o acesso a dados, serviços e conteúdo por meio de interações com o governo.* ”.

Estando caracterizada esta interdependência de diversos atores apresenta-se o cenário de recomendações das possíveis abordagens em projetos de eGov que auxiliem a resolver as questões de interoperação e heterogeneidade. No relatório do Governo Brasileiro intitulado “EGD – Estratégia de Governança Digital” (MP, 2016, p.26) estas questões são tratadas como princípios estratégicos, assim endereçadas:

- **Interoperação** relaciona-se com o princípio número 3 – Compartilhamento da capacidade de serviço – definido como: “... *órgãos e entidades deverão compartilhar infraestrutura, sistemas e serviços, de forma a evitar duplicação de*

¹ Valor público: produtos e resultados gerados, preservados ou entregues pelas atividades de uma organização que representem respostas efetivas e úteis às necessidades ou às demandas de interesse público e modifiquem aspectos do conjunto da sociedade ou de alguns grupos específicos reconhecidos como destinatários legítimos de bens e serviços públicos. (MP, 2016a)

² No âmbito do Governo Brasileiro o conceito foi expandido para o de Governança Digital: “... segundo o qual o cidadão torna-se partícipe da construção de políticas públicas, que já nascem em plataformas digitais.” (MP, 2016a).

esforços, eliminar desperdícios e custos e reduzir a fragmentação da informação em silos. ”

- **Heterogeneidade** relaciona-se com o princípio número 4 – Compartilhamento de dados – definido como: “ ... *órgãos e entidade da administração pública federal direta, autárquica e fundacional deverão compartilhar dados entre si, sempre que houver oportunidade de simplificar processos administrativos e a prestação de serviços à sociedade. ”*

Sobre **Interoperação** em projetos de eGov utiliza-se duas referências:

- O ePing (MP, 2018) em todas as suas áreas de recomendações: interconexão, segurança, meios de acesso, organização e intercâmbio de informações, áreas de integração para governo eletrônico.
- O FACIN (MP, 2017a) que *"tem por objetivo construir uma visão consistente dos modelos de representação das capacidades de negócio de cada organização governamental."* Em seu Modelo de Referência (MP, 2017c) o FACIN apresenta recomendações para promover o estabelecimento de práticas uniformes de governança nas organizações governamentais brasileiras. E, em seu Modelo de Conteúdo (MP, 2017b), descreve-se a dimensão semântica da interoperabilidade, com vistas a “ ... *garantir que os dados trocados tenham seu significado corretamente interpretado dentro do contexto de uma determinada transação ou busca de informação ... ”.*

Sobre **Heterogeneidade** em projetos de eGov adota-se como referência a Parte II, capítulo 5, do ePing: Áreas de Integração para Governo Eletrônico (MP, 2018, p. 33). Na sua Tabela 15 – Especificações para Organização e Intercâmbio de Informações: Vocabulários e Ontologias – constam recomendações para: descrição de recursos, sintaxe RDF, especificação de vocabulários para RDF, vocabulários, sistemas de organização do conhecimento, linguagem de definição de ontologias na web e linguagem de consulta semântica.

Neste cenário de recomendações e esforços de padronização, apresenta-se o estudo dos Vocabulários Controlados, como artefato fundamental para a organização do conhecimento. A necessidade deste estudo está explícita em (MP, 2017b) quando faz referência às taxonomias de cada uma das suas 9 visões, que devem prover *"um conjunto de termos de classificação e uma terminologia unificada dos elementos de conceitos selecionados para cada Visão, dentro do contexto do Governo como um todo."*

Taxonomias são estruturas que se organizam com a classificação e a criação de categorias. De modo complementar, ontologia se refere a um artefato constituído por um vocabulário usado para descrever uma certa realidade, mais um conjunto de fatos explícitos e aceitos que dizem respeito ao sentido pretendido para as palavras do vocabulário (ALMEIDA; BAX, 2003). Temos aqui então referenciada a necessidade de um vocabulário de conceitos e termos, neste caso, referentes ao domínio "governo brasileiro". Trata-se de um vocabulário controlado.

Conforme (MACULAN, 2011) *"vocabulário controlado é uma lista de termos elaborada para identificar o assunto ou os assuntos de um documento com especificidade bastante para permitir sua recuperação rápida e eficaz."* Desta forma, um vocabulário controlado possibilita o controle da terminologia de um domínio, instituindo regras para seu uso, compartilhamento e expansão.

No ePing constam como recomendações para Vocabulários Controlados a Lista de vocabulários recomendados pela W3C³ e a Lista de vocabulários recomendados pelo governo⁴.

No Repositório de Vocabulários e Ontologias do Governo Eletrônico recomendam-se diversos vocabulários. O que nos interessa no momento é o VCGE – Vocabulário Controlado do Governo Eletrônico – por ser o mais genérico, tendo assim a “vocação” de ser o vocabulário de domínio para assuntos de governo.

O VCGE é apresentado por um documento bastante completo, tanto técnica quanto conceitualmente (MP, 2016b). Note-se, especialmente o capítulo 2 – Construção do VCGE. Este capítulo é composto de 2 subitens: Sistema Conceitual e Metodologia para Elaboração do VCGE. Na nossa visão, estas bases conceituais e metodológicas continuam válidas para quaisquer novas modelagens e revisões que venham a ser feitas.

O ponto inicial da nossa avaliação é a Tese de Mestrado de João P. R. Ramirez (RAMIREZ, 2015). O objetivo do trabalho em referência é analisar a evolução do VCGE segundo seus objetivos e critérios aplicáveis a taxonomias e tesouros. Com base na bibliografia da Ciência da Informação Ramirez utiliza-se de critérios, avaliando detalhadamente cada um deles. Transcreve-se a conclusão da análise: “... a versão 2.0.3 (do VCGE) atende totalmente a 5 e parcialmente a 3 critérios e não atende a 2 critérios. Os resultados demonstram que ainda são necessários ajustes para que o vocabulário atenda plenamente aos critérios empregados na literatura da área de Ciência da Informação.”

Sendo assim, coloca-se o passo inicial da pesquisa, que é analisar as alternativas de evolução do VCGE.

3. Metodologia

O trabalho em andamento é a realização uma pesquisa aplicada – aplicação dos conhecimentos básicos na geração de novos produtos, processos e serviços – com objetivos exploratórios – descoberta de teorias e práticas que modificarão as existentes – empregando procedimentos de pesquisa experimental – descoberta de novos materiais, métodos, técnicas, protótipos de software (JUNG, 2004).

O trabalho iniciou-se em agosto de 2016 e tem término previsto para julho de 2020. No primeiro semestre de 2018 – data atual – o projeto está na fase de Revisão de Literatura e estabelecimento do marco teórico.

Como objetivos específicos tem-se:

- Propor alternativas de caminhos evolutivos para o VCGE.
- Analisar a disponibilidade de Vocabulários e Taxonomias de eGov em formato de utilização pública e sua possibilidade de reuso pelo Governo Brasileiro.
- Propor alternativas construtivas aos projetos de disponibilização de dados do tipo “data-lake” através de Modelagem Semântica (ALSERAFI *et al.*, 2017; FANG, 2015; MILOSLAVSKAYA; TOLSTOY, 2016).

³ Como definido em <http://www.w3.org/2011/rdfa-context/rdfa-1.1> onde podem ser encontradas referências para diversos vocabulários padronizados e de código aberto, disponíveis para reuso.

⁴ Como definido em <http://vocab.e.gov.br/>

- Complementar os projetos de Arquitetura Corporativa através da Análise Semântica (AZEVEDO *et al.*, 2015; BAKAR; HARIHODIN; KAMA, 2016; GRIFFO *et al.*, 2017; GUARINO; GUIZZARDI; SALES, 2018; ROUHANI *et al.*, 2015).
- Avaliar os possíveis ganhos dos mecanismos de eGov através da melhor compreensão dos modelos de dados em diversos órgãos e apontando possíveis melhoras em processos e sistemas atualmente em uso.
- Propor mecanismos para a melhora da Governança Pública através da elaboração de mecanismos de integração, visualização e compreensão dos Dados Governamentais Públicos, permitindo maior transparência aos processos e resultados de governo.

4. Considerações Finais

Como resultado inicial, parece ser um promissor candidato a reuso o Eurovoc, que é um tesouro multilíngue e pluridisciplinar sobre as atividades da União Europeia que contém termos em 26 das suas línguas⁵. O Eurovoc apresenta diversas características que o tornam interessante como possibilidade de uso pelo ePing e, conseqüentemente, pelo FACIN. Dentre estas destacamos que nos parece ser decisivo o fato de apresentar uma versão em português⁶ e estar disponível para download e uso público. Podemos citar outras características positivas adicionais: tem ampla aceitação mundial, possui um mecanismo de atualização em funcionamento e pode servir como padrão para diversas outras iniciativas de eGov, não somente no FACIN e no ePing.

Bibliografia

ALMEIDA, Maurício Barcellos; BAX, Marcello Peixoto. Taxonomia para projetos de integração de fontes de dados baseados em ontologias. 2003, [S.l: s.n.], 2003. p.20.

ALSERAFI, Ayman *et al.* Towards Information Profiling: Data Lake Content Metadata Management. 2017, [S.l: s.n.], 2017. 9781509054725. .

AZEVEDO, Carlos L.B. *et al.* Modeling resources and capabilities in enterprise architecture: A well-founded ontology-based proposal for ArchiMate. **Information Systems** v. 54, p. 235–262 , 2015.978-0-7695-5081-7.

BAKAR, Nur Azaliah A; HARIHODIN, S; KAMA, Nazri. Assessment of Enterprise Architecture Implementation Capability and Priority in Public Sector Agency. **Procedia Computer Science** International Conference on Enterprise Information Systems, v. 100, p. 198–206 , 2016.

FANG, Huang. Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem. jun. 2015, [S.l.]: IEEE, jun. 2015. p.820–824.

⁵ "O Eurovoc é gerido pelo Serviço das Publicações, que passou a aplicar uma gestão de tesouros baseada em ontologias e tecnologias da Web semântica de acordo com as recomendações do consórcio World Wide Web (W3C) e com as últimas tendências em matéria de normas aplicáveis aos tesouros. Entre os utilizadores do Eurovoc encontram-se as instituições da União Europeia, o Serviço das Publicações da UE, os parlamentos nacionais e regionais da Europa, as administrações nacionais e utilizadores privados em todo o mundo." <http://eurovoc.europa.eu/drupal/?q=pt>

⁶ Trata-se do português de Portugal, mas isto não nos parece representar uma dificuldade muito grande.

GRIFFO, C *et al.* From an Ontology of Service Contracts to Contract Modeling in Enterprise Architecture. 2017, [S.l: s.n.], 2017. p.40–49. VO - .

GUARINO, Nicola; GUIZZARDI, Giancarlo; SALES, Tiago Prince. On the Ontological Nature of REA Core Relations. 2018, [S.l: s.n.], 2018. p.9.

JUNG, Carlos Fernando. **Metodologia para pesquisa & desenvolvimento: aplicada a novas tecnologias, produtos e processos**. 1. ed. São Paulo: [s.n.], 2004. 312 p. .

MACULAN, Benildes Coura M. S. **Taxonomia Facetada Navegacional: construção a partir de uma matriz categorial para trabalhos acadêmicos**. UFMG, 2011. 195 p.

MILOSLAVSKAYA, Natalia; TOLSTOY, Alexander. Big Data, Fast Data and Data Lake Concepts. 2016, [S.l: s.n.], 2016. 978-1-5090-3946-3. .

MP. Ministério do Planejamento Desenvolvimento e Gestão. **ePing - Padrões de Interoperabilidade do Governo Eletrônico**. Brasília: [s.n.], 2018. Disponível em: <<https://www.governodigital.gov.br/transformacao/orientacoes/interoperabilidade/eping-padroes-de-interoperabilidade-de-governo-eletronico>>.

MP. Ministério do Planejamento Desenvolvimento e Gestão. **Estratégia de Governança Digital - EGD**. Brasília: [s.n.], 2016a. Disponível em: <http://www.planejamento.gov.br/seminariodigital/seminario/egd_cartilha_capa_miolo_final_avulsascorrecao.pdf>.

MP. Ministério do Planejamento Desenvolvimento e Gestão. **FACIN - Framework de Arquitetura Corporativa para Interoperabilidade no Apoio à Governança**. Brasília: [s.n.], 2017a. Disponível em: <<https://www.governodigital.gov.br/transformacao/orientacoes/interoperabilidade/governanca-e-interoperabilidade/facin>>.

MP. Ministério do Planejamento Desenvolvimento e Gestão. **FACIN - Modelo de Conteúdo**. Brasília: [s.n.], 2017b.

MP. Ministério do Planejamento Desenvolvimento e Gestão. **FACIN - Modelo de Referência**. Brasília: [s.n.], 2017c.

MP. Ministério do Planejamento Desenvolvimento e Gestão. **VCGE - Vocabulário Controlado do Governo Eletrônico**. Brasília: [s.n.], 2016b. Disponível em: <<https://www.governodigital.gov.br/transformacao/orientacoes/interoperabilidade/vocabulario-controlado-do-governo-eletronico-vcge>>.

OECD. Organisation for Economic Co-operation and Development. **Recommendation of the Council on Digital Government Strategies**. [S.l: s.n.], 2014.

RAMIREZ, João P. Righi. **Vocabulário Controlado do Governo Eletrônico (VCGE): Uma análise com base em em critérios aplicáveis a taxonomias e tesouros**. Universidade Federal de Santa Catarina, 2015. 189 p.

ROUHANI, Babak Darvish *et al.* A systematic literature review on Enterprise Architecture Implementation Methodologies. **Information and Software Technology** v. 62, p. 1–20 , 2015.

UN. United Nations Dept of Economic and Social Affairs. **United Nations - E-Government Survey 2016**. [S.l: s.n.], 2016. Disponível em: <<http://workspace.unpan.org/sites/Internet/Documents/UNPAN97453.pdf>>.