



Transinformação

ISSN: 0103-3786

transinfo@puc-campinas.edu.br

Pontifícia Universidade Católica de
Campinas
Brasil

Silva RIBEIRO, Cláudio José; Vieira PEREIRA, Durval
A publicação de dados governamentais abertos: proposta de revisão da classe sobre
Previdência Social do Vocabulário Controlado do Governo Eletrônico
Transinformação, vol. 27, núm. 1, enero-abril, 2015, pp. 73-82
Pontifícia Universidade Católica de Campinas
Campinas, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=384351519008>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica
Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

A publicação de dados governamentais abertos: proposta de revisão da classe sobre Previdência Social do Vocabulário Controlado do Governo Eletrônico

The publication of open government data: A proposal for the revision of class in Social Security Regarding the Controlled Vocabulary of the Electronic Government

Cláudio José Silva RIBEIRO¹

Durval Vieira PEREIRA²

Resumo

Este trabalho analisa a classe sobre Previdência Social do Vocabulário Controlado do Governo Eletrônico, investigando a publicação de dados governamentais abertos, realizada pela Empresa de Tecnologia e Informações da Previdência Social. Trata-se de um estudo exploratório, com objetivo de propor a revisão da classe sobre Previdência Social no domínio do Governo Eletrônico brasileiro, a fim de ampliar e corrigir a representação da área. O trabalho utiliza como metodologia: (1) pesquisa bibliográfica, para fundamentação teórica; (2) seleção da classe de assunto sobre Previdência Social do Vocabulário Controlado do Governo Eletrônico, como recorte; (3) análise dos pontos relacionados à categorização, linguagem escolhida (natural ou artificial), cobertura dos termos pertencentes ao escopo da classe de assunto estudado e relações apresentadas entre os termos; (4) propostas para revisão da classe estudada. O estudo aponta, nos resultados, falhas estruturais e inconsistências conceituais localizados no vocabulário. Discute a necessidade de se utilizarem vocabulários consistentes para representação dos dados publicados, a fim de torná-los legíveis por máquina e compartilháveis entre aplicações de *software*. Afirma que se deve investir em vocabulários já existentes para evitar esforços com a criação de outros instrumentos de representação. Julga necessário que outros órgãos públicos, além da Empresa de Tecnologia e Informações da Previdência Social, interessem-se em analisar e propor alterações e atualizações nas classes cuja temática contemple os dados por eles geridos. Conclui que a utilização do Vocabulário Controlado do Governo Eletrônico pode ir além de um mapa para navegação em páginas eletrônicas do setor público brasileiro, constituindo instrumento dinâmico de organização do conhecimento, podendo ser utilizado para diversos fins, de acordo com os interesses de cada usuário.

Palavras-chave: Dados governamentais abertos. Previdência social. Taxonomia. Vocabulário controlado.

Abstract

The aim of the paper is to examine the controlled vocabulary class on Social Security of the Electronic Government, investigating the publication of open government data held by the Empresa de Tecnologia e Informações da Previdência Social. This is an exploratory study with the purpose of proposing a vocabulary class revision on Social Security in the domain of the Brazilian Electronic Government, in order to expand and correct representation of the area. The methodology used was as follows: (1) literature research and review; (2)

¹ Universidade Federal do Estado do Rio de Janeiro, Departamento de Processos Técnicos e Documentais, Programa de Pós-Graduação em Biblioteconomia. Av. Pasteur, 458, Sala 413, Urca, 22290-240, Rio de Janeiro, RJ, Brasil. Correspondência para/Correspondence to: C.J.S. RIBEIRO. E-mail: <claudio.ribeiro@unirio.br>.

² Universidade Federal Fluminense, Instituto de Arte e Comunicação Social, Programa de Pós-Graduação em Ciência da Informação. Rio de Janeiro, RJ, Brasil. Recebido em 15/5/2013, reapresentado em 27/11/2013 e aceito para publicação em 17/1/2014.

selection of the thematic class that will be studied; (3) analysis of the aspects related to categorization, use of language (natural or artificial); coverage of terms belonging to the scope of the vocabulary class studied, and relationships between the terms; (4) suggestions for reviewing social security class. The study identifies the structural weakness and conceptual inconsistencies found in the vocabulary. We further discuss the need to use consistent vocabulary representation for the data published with the intent of making them readable by machine and shareable among software applications. We affirm that there must be investments in existing vocabularies rather than efforts to develop new representation instruments. We think it is important that other public bodies, other than Empresa de Tecnologia e Informações da Previdência Social, become interested in analyzing and proposing changes and updates on classes whose theme contemplates the data managed by them. We conclude that the use of the Controlled Vocabulary of Electronic Government can go beyond a map for surfing web pages of the Brazilian public sector, consisting of a dynamic tool of knowledge organization that can be used for various purposes, in accordance with the interests of each user.

Keywords: Open government data. Social security system. Taxonomy. Controlled vocabulary.

Introdução

Atualmente, são investidos diversos recursos para a melhoria da comunicação entre o Governo e os cidadãos. Cada vez mais as tecnologias de informação estão sendo utilizadas para aprimorar a qualidade dos produtos e serviços oferecidos pelo Estado à população, que busca uma maior participação nas práticas de gestão pública. A principal iniciativa utilizada pelo Estado para efetivar esta interação entre os cidadãos e o poder público tem sido o Governo Eletrônico.

Baseados na “Proposta de Política de Governo Eletrônico”, Chahin *et al.* (2004) esclarecem que o Governo Eletrônico surge no ano de 2000 com a proposta de estabelecer um novo paradigma cultural de inclusão digital, centrado no cidadão, com a redução de custos unitários, a melhoria da gestão e qualidade dos serviços públicos, a transparência e a simplificação de processos. Os autores complementam que, àquela época, eram fundamentais a inclusão social e a ampliação do uso da tecnologia de informação pelo cidadão, o que possibilitaria a universalização e a democratização do acesso aos serviços.

O projeto de Governo Eletrônico pautava-se, inicialmente, no acesso a equipamentos e à Internet por meio da instalação de telecentros, uma vez que esta era ainda restrita a alguns grupos sociais. Apesar disso, já era possível perceber iniciativas governamentais para fomentar a ampliação do uso da tecnologia da informação e comunicação pelo cidadão (Neves Júnior, 2006).

Assim, desde o princípio dos anos 2000 até hoje, verifica-se que muitas pesquisas no âmbito do Governo

contribuíram para o desenvolvimento de metodologias, técnicas e equipamentos necessários para a modernização da administração pública. Essas pesquisas foram desenvolvidas com o intuito de aprimorar os processos operacionais e administrativos, que passaram a ser associados com a prestação de serviços públicos eletrônicos (Ribeiro & Almeida, 2011).

Uma dessas iniciativas foi voltada para os dados governamentais abertos, o que pode ser entendido como a “oferta de bases de dados em estado bruto, para serem livremente manipuladas, filtradas ou cruzadas com outras, inclusive permitindo a construção de novas aplicações e conhecimentos pela própria sociedade civil” (Vaz *et al.*, 2010, p.46). Dessa forma, a publicação dos dados governamentais abertos utiliza-se das tecnologias da Web Semântica para que seu uso proporcione melhorias na acessibilidade e reutilização dos dados.

Ademais, para que esse propósito seja atingido, é imprescindível a utilização de vocabulários para descrição dos dados governamentais. Os vocabulários podem ser utilizados de duas maneiras:

[...] como esquemas, fornecendo um conjunto de propriedades que um objeto possa ter, e como esquemas de codificação de vocabulário, descrevendo a faixa de valores que uma determinada propriedade pode tomar, em termos formais. A utilização de vocabulários interligados ajuda na aquisição de conhecimento por controlar rigorosamente e contextualizar os dados (conceitos, objetos, etc.). Essa abordagem permite que os processos de metadados familiares, enquanto a capacidade de vincular os dois vocabulários e registros de dados fornece uma infraestrutura que permite a descoberta de informação mais

eficaz e uso (Méndez & Greenberg, 2012, p.237, tradução nossa)³.

A Empresa de Tecnologia e Informações da Previdência Social (Dataprev) destinou seus esforços à publicação de dados no âmbito do Ministério da Previdência Social (MPS). O primeiro projeto teve seu escopo dedicado às informações sobre acidentes de trabalho, coletados dos dados descritos no Anuário Estatístico de Acidentes de Trabalho (AEAT). Por meio dessas informações, é possível, por exemplo, construir um diagnóstico mais preciso acerca da epidemiologia desses acidentes e propiciar a elaboração de políticas mais eficazes para as áreas relacionadas (Ribeiro & Almeida, 2011).

Para que os princípios de dados abertos sejam completamente atendidos, é preciso que haja uma organização semântica dos dados publicados. Assim, pesquisando instrumentos de representação semântica de informações previdenciárias existentes, os autores deste estudo localizaram o Vocabulário Controlado do Governo Eletrônico (VCGE). Após uma análise sobre a classe do assunto Previdência Social, dentro do domínio do Governo Eletrônico brasileiro, chegou-se à conclusão de que o domínio representado poderia ser revisto para ampliar a abrangência de termos e o universo de possíveis usuários.

Dessa forma, este estudo tem o objetivo de propor uma revisão da classe de Previdência Social representada no VCGE, a fim de ampliar e corrigir a representação atualmente disponibilizada, abrindo possíveis caminhos para o uso da semântica na publicação de dados abertos pela Dataprev.

Ainda hoje, a maioria das informações divulgadas pelo Governo está publicada em formatos proprietários ou com tecnologia que restringe seu acesso. Apenas publicar os dados não os torna acessíveis. É necessário que se permita desagregar, reprocessar e reutilizar tais dados, buscando novas aplicações e utilidades.

Alguns países já adotam a iniciativa de dados governamentais abertos, como, por exemplo, Estados

Unidos da América <www.data.gov>, Alemanha <www.offenedatten.de>, Grã-Bretanha <data.gov.uk>, Nova Zelândia <data.govt.nz> e Austrália <data.gov.au>, dentre outros. Os resultados obtidos são cidadãos que aproveitam dados do governo para criar serviços de interesse da população e até mesmo de ajuda ao governo, como mapa de áreas com risco de enchentes, localização de imóveis de acordo com características indicadas, informações sobre áreas com segurança ou com criminalidade alta, planejamento de gastos com energia elétrica, além de outros serviços que auxiliam a população, sem dispêndio de recursos públicos para tais ações.

O *World Web Consortium* Brasil (2009, p.39) define dados governamentais abertos como a publicação de informações do setor público “em formato bruto aberto, de maneira a torná-las acessíveis a todos e permitir sua reutilização, como a criação de *mashups* de dados”. Ademais, os dados governamentais abertos propiciam “a análise destes dados segundo vários pontos de vista, [e] esta disseminação permite que a sociedade construa uma visão mais clara sobre o desempenho do governo frente às metas estabelecidas, bem como sobre o desenvolvimento de políticas públicas” (Ribeiro & Almeida, 2011, p.2571).

Dessa forma, pode-se dizer que os dados governamentais abertos buscam a publicação na *Web* de informações do setor público em formato bruto e aberto, legíveis não somente por pessoas, mas também compreensíveis por máquinas, de modo a permitir sua reutilização em novas aplicações digitais desenvolvidas pela sociedade.

Bernes-Lee (2009, p.2, tradução nossa)⁴ conclui que os dados governamentais devem ser publicados de forma aberta por três motivos: “aumentar a consciência cidadã das funções do governo para permitir maior responsabilidade; contribuir com informações valiosas sobre o mundo; e permitir que o governo, o país e o mundo funcionem com mais eficiência”.

³ Texto original: “Linked data uses vocabularies in two ways: as schemas, providing a set of properties that an object might have; and as vocabulary encoding schemes describing the range of values a given property might take, in informal terms. Linked vocabularies aid knowledge acquisition by tightly controlling and contextualizing data (concepts, objects, etc.). This approach enables familiar metadata processes, while the ability to link both vocabularies and data records themselves provides an infrastructure that allows more effective information discovery and use” (Méndez & Greenberg, 2012, p.237).

⁴ Texto original: “Government data is put online typically for 3 reasons: 1) Increasing citizen awareness of government functions to enable greater accountability; 2) Contributing valuable information about the world; and 3) Enabling the government, the country, and the world to function more efficiently” (Bernes-Lee, 2009, p.2).

O *Open Government Data* (2007) desenvolveu os oito princípios dos dados governamentais abertos, que nortearam diversos trabalhos posteriores sobre o tema, como os de Vaz *et al.* (2010, p.47) e *World Web Consortium Brasil* (2011, p.14).

1. *Completo*. Todos os dados públicos estão disponíveis. Dado público é o dado que não está sujeito a limitações válidas de privacidade, segurança ou controle de acesso.
2. *Primários*. Os dados são apresentados tais como coletados na fonte, com o maior nível de granularidade e sem agregação ou modificação.
3. *Atuais*. Os dados são disponibilizados tão rapidamente quanto necessário à preservação do seu valor.
4. *Acessíveis*. Os dados são disponibilizados para o maior alcance possível de usuários e para o maior conjunto possível de finalidades.
5. *Compreensíveis por máquinas*. Os dados são razoavelmente estruturados de modo a possibilitar processamento automatizado.
6. *Não discriminatórios*. Os dados são disponíveis para todos, sem exigência de requerimento ou cadastro.
7. *Não proprietários*. Os dados são disponíveis em formato sobre o qual nenhuma entidade detenha controle exclusivo.
8. *Livres de licença*. Os dados não estão sujeitos a nenhuma restrição de direito autoral, patente, propriedade intelectual ou segredo industrial. São permitidas restrições sensatas relacionadas à privacidade, segurança e privilégios de acesso (Vaz *et al.*, 2010, p.47).

A necessidade de facilitar a compreensão dos dados, tanto por pessoas quanto por máquinas, leva à *Web Semântica* como plataforma para melhorar o acesso e possibilitar o processamento automatizado por máquinas, bem como para viabilizar buscas que obtenham resultados mais precisos (Bernes-Lee *et al.*, 2001). Ao fornecer significados às “coisas” na *Web*, de modo que a linguagem seja compreendida por máquina, capturando, processando e relacionando dados de diferentes fontes, a *Web Semântica* viabiliza que a publicação dos dados governamentais abertos seja encontrada por quem a reutilize, dando outra utilidade àqueles dados. Bernes-Lee (2009) convalida essa questão quando afirma que a publicação de dados governamentais abertos deve se

basear na *Web Semântica*, pois assim os dados e as conexões com documentos receberão significados definidos, com metadados em formato padronizado, possibilitando a compreensão de pessoas e computadores.

Dessa forma, conforme Souza e Alvarenga (2004), a *Web Semântica* pretende também contribuir para o reuso da informação, facilitando sua localização, acesso, comunicação e interação entre diferentes fontes.

Nesse caso, a *Web Semântica* utiliza para essa interação a abordagem de *Linked Open Data* (LOD), noção desenvolvida por Berners-Lee (2006), que tem o objetivo de ligar dados de diferentes fontes, criando uma rede de compartilhamento de informações. Essa rede é viabilizada pelo uso de padrões, bem como pela elaboração de vocabulários estruturados, que possibilitem a padronização terminológica, visando à comunicação e compreensão dos dados descritos por máquinas.

Sobre esse aspecto, Berners-Lee *et al.* (2001, p.6) esclarecem que o desafio da *Web Semântica* é “fornecer uma linguagem que expresse tanto os dados quanto as regras para o raciocínio sobre os dados e que permita que as regras de qualquer sistema de representação do conhecimento existente possam ser exportadas para a *Web*”.

Adicionalmente, é lícito supor que as teorias provenientes da Organização do Conhecimento podem contribuir para dar suporte à padronização linguística necessária na *Web Semântica*. Uma vez que esse campo do conhecimento é responsável por importantes estudos sobre vocabulários estruturados e formalizados, tais como as taxonomias e as ontologias, tais instrumentos de representação do conhecimento podem possibilitar que a *Web* trabalhe com os conceitos e suas relações, para uma eficaz recuperação e utilização da informação.

De acordo com a área do saber que estuda o conceito de Taxonomia, este recebe uma definição própria para aquela área. No âmbito da Ciência da Informação, o termo taxonomia é

[...] utilizado para identificar as classificações baseadas nas hierarquias de assuntos, atuar como um instrumento capaz de organizar e recuperar as informações, com o auxílio de sinônimos e suas relações hierárquicas e associativas, além de garantir o controle do vocabulário específico da área (Aganette *et al.* 2010, p.6).

Sob o prisma da *Web Semântica*, o uso de taxonomias “permite que se estabeleçam padrões de alto nível para a ordenação e a classificação de informação através do uso de mecanismos de herança” (Campos & Gomes, 2008, p.2).

Ainda na perspectiva de instrumentos de representação do conhecimento para utilização na *Web Semântica*, a ontologia é apresentada como “um vocabulário comum para uma comunidade que precisa compartilhar informação em um determinado domínio. Inclui definições de conceitos básicos no domínio e as relações entre estes, de forma que sejam interpretáveis por máquina” (Marcondes & Campos, 2008, p.121).

Alguns estudos que envolvem *Web Semântica* denominam taxonomia como uma espécie de “ontologia simples” (McGuinness, 2003). Já outros pesquisadores conseguem observar uma diferença entre taxonomia e ontologia, pois:

Enquanto as taxonomias buscam o desenvolvimento de categorias para facilitar a inserção e recuperação da informação, as ontologias vão além, objetivando o desenvolvimento de um ‘consenso linguístico’ em áreas específicas. Ontologias representam o conhecimento de um dado domínio em forma de uma rede relacional, intensional, onde as relações se sobrepõem aos possíveis ‘estados da coisa’. As ontologias podem ser consideradas mais complexas que as taxonomias, levando em consideração que não se limitam aos relacionamentos taxonômicos adotados na ordenação de classes e subclasses, estabelecendo outros tipos de relações semânticas, como as de associação, derivadas da explicitação das características dos conceitos (Vital & Café, 2011, p.126).

Dessa forma, os vocabulários controlados, dentre os quais destacam-se as Taxonomias e as Ontologias, podem contribuir para estabelecer relações semânticas, entendidas, no caso da *Web Semântica*, não somente por homens, mas também por agentes de *softwares*. Além disso, cooperam para a interoperabilidade, a detecção de inconsistências, o formalismo para representação e a reutilização de conhecimentos.

Percebe-se que existe uma proximidade, quase que dependente, entre dados abertos, *Web Semântica* e instrumentos de Organização do Conhecimento, como ontologias e taxonomias. Na verdade, há áreas do conhe-

cimento que estão lidando com temáticas em comum, para as quais a parceria é o fator que acarreta o desenvolvimento recíproco.

Métodos

O recorte deste estudo contempla a taxonomia como arcabouço de sustentação para o desenvolvimento de ontologias na Previdência Social.

Tendo em vista tal recorte, o estudo tem caráter exploratório, na medida em que busca identificar pontos, na classe de Previdência Social do VCGE, que poderiam ser melhorados. Dessa forma, a metodologia foi dividida em etapas: 1) pesquisa bibliográfica que fornecesse o referencial teórico necessário para verificar como determinadas questões conceituais e estruturais são apresentadas pelo VCGE; 2) seleção da classe de assunto sobre Previdência Social do VCGE, no domínio do Governo Eletrônico brasileiro, como recorte a ser observado; 3) identificação de possíveis inconsistências relacionadas à categorização, especialmente sobre: linguagem escolhida (natural ou artificial); cobertura dos termos pertencentes ao escopo da classe de assunto estudado; e relações apresentadas entre os termos; 4) propostas para revisão/alteração da classe de assunto sobre Previdência Social.

O caso aqui apresentado é restrito e ilustrativo, não exaustivo. A pesquisa teve como propósito levantar a discussão sobre a utilização do VCGE como instrumento de organização do conhecimento para a publicação de dados governamentais abertos.

Resultados

O Vocabulário Controlado do Governo Eletrônico é uma evolução da Lista de Assuntos do Governo (LAG). É um esquema para ser utilizado no elemento assunto categoria (*subject.category*) do Padrão de Metadados do Governo Eletrônico (e-PMG). Para os gerentes de sítios e portais governamentais, os termos retirados do VCGE tornarão mais direta a apresentação dos serviços, disponibilizados em uma estrutura de diretório baseada em indexadores (Brasil, 2011).

O Vocabulário estudado apresenta as relações genéricas e específicas entre os termos, caracterizando

uma taxonomia. De acordo com Moreiro González (2011, p.51), a taxonomia “serve para ordenar informação em uma hierarquia [...] em forma de árvore que estabelece relação entre os objetos de generalização-especialização”. Apesar do Brasil (2011) afirmar que em seu processo de construção utilizaram os princípios de um tesouro, o VCGE se aproxima mais de uma taxonomia do que de um tesouro, pois este último abrange uma quantidade maior de relações (Gomes, 1990), como, por exemplo, a partitiva e a associativa, não utilizadas no VCGE.

De acordo com o Brasil (2011), o público-alvo do VCGE é o cidadão. O esquema busca ser intuitivo para o cidadão que busque, nas páginas governamentais *online*, assuntos do seu interesse. Nesse sentido, o VCGE descreve sua metodologia nos seguintes termos:

- 1) prefere a linguagem do leigo ao jargão do serviço público e aos termos técnicos;
- 2) não supõe que o cidadão tenha conhecimento prévio das responsabilidades de cada nível ou órgão governamental. Procura ser independente da estrutura governamental, devendo sobreviver às mudanças de estruturas e organogramas;
- 3) o uso comum é mais importante do que a precisão acadêmica, quando se está escolhendo nomes ou posições relativas aos cabeçalhos (Brasil, 2011, *online*).

Então, pode-se inferir que o VCGE utilizou a “linguagem do leigo”, ou seja, uma linguagem livre e preocupada com o uso comum mais do que com a “precisão acadêmica”. No entanto, surge aqui uma primeira inquietação: a precisão de vocabulários controlados desenvolvida em estudos acadêmicos não vem justamente para fornecer teorias e métodos para uma maior precisão na descrição dos termos e, consequentemente, melhorar a recuperação da informação?

Vislumbra-se que a utilização do Vocabulário pode ir além de um mapa para navegação em páginas eletrônicas do setor público brasileiro. Seu uso pode ser estendido a outros propósitos, como, por exemplo, ser utilizado pelo Estado para representação do domínio - mapa conceitual -, de cada setor público, mostrando sua trajetória ao longo do tempo, ou pode ser usada pelos cidadãos para descrição, recuperação e reutilização de informações governamentais, atendendo assim a organização necessária para a publicação dos dados abertos.

Ao analisar a classe referente à Previdência Social esquematizada no VCGE (Figura 1), percebe-se que a representação feita da respectiva área não corresponde a sua abrangência temática. Termos importantes para entendimento da Previdência Social não estavam contemplados no Vocabulário.

Um exemplo dessa falta de cobertura é a ausência do termo “pensão”: se um cidadão necessitar buscar a espécie de pensão que o beneficie, não a encontrará, pois esse termo não é classificado pelo VCGE. Estão ausentes da classe do VCGE sobre Previdência Social as pensões oferecidas pelo Ministério da Previdência Social (Brasil, 2012a): pensão especial a pessoas atingidas pela hanseníase; pensão especial aos dependentes de vítimas fatais por contaminação na hemodiálise Caruaru (PE); pensão especial mensal vitalícia; pensão especial vitalícia; pensão mensal vitalícia do dependente do seringueiro; pensão mensal vitalícia do seringueiro; pensão mensal vitalícia por síndrome de talidomida; pensão por morte (Ex-Sasse); pensão por morte de ex-combatente; pensão por morte por acidente de trabalho; pensão previdenciária por morte.

Outro ponto relevante é que a Previdência Social está subordinada à classe de “Pessoas, família e sociedade”. Entretanto, outras áreas como Saúde, Trabalho e Educação, apesar de estarem conceitualmente ligadas aos direitos e benefícios da pessoa e da sociedade, encontram-se em classes independentes. De acordo com Campos e Gomes (2008, p.5), a “categorização é um processo que requer pensar o domínio de forma dedutiva, ou seja, determinar as classes de maior abrangência dentro da temática escolhida”.

Dessa maneira, não ficaram claros os princípios utilizados para o processo da categorização da classe de assunto da Previdência Social dentro do domínio do Governo Eletrônico. A estrutura organizacional do Governo Eletrônico classifica, no mesmo nível hierárquico, Previdência Social, Saúde, Trabalho e Educação dentro de Ministérios (Brasil, 2013b). Assim, poderia ser revista a categorização aplicada no VCGE, de forma a “analisar o domínio a partir de recortes conceituais que permitem determinar a identidade dos conceitos (categorias) que fazem parte deste domínio” (Campos & Gomes, 2008, p.5).

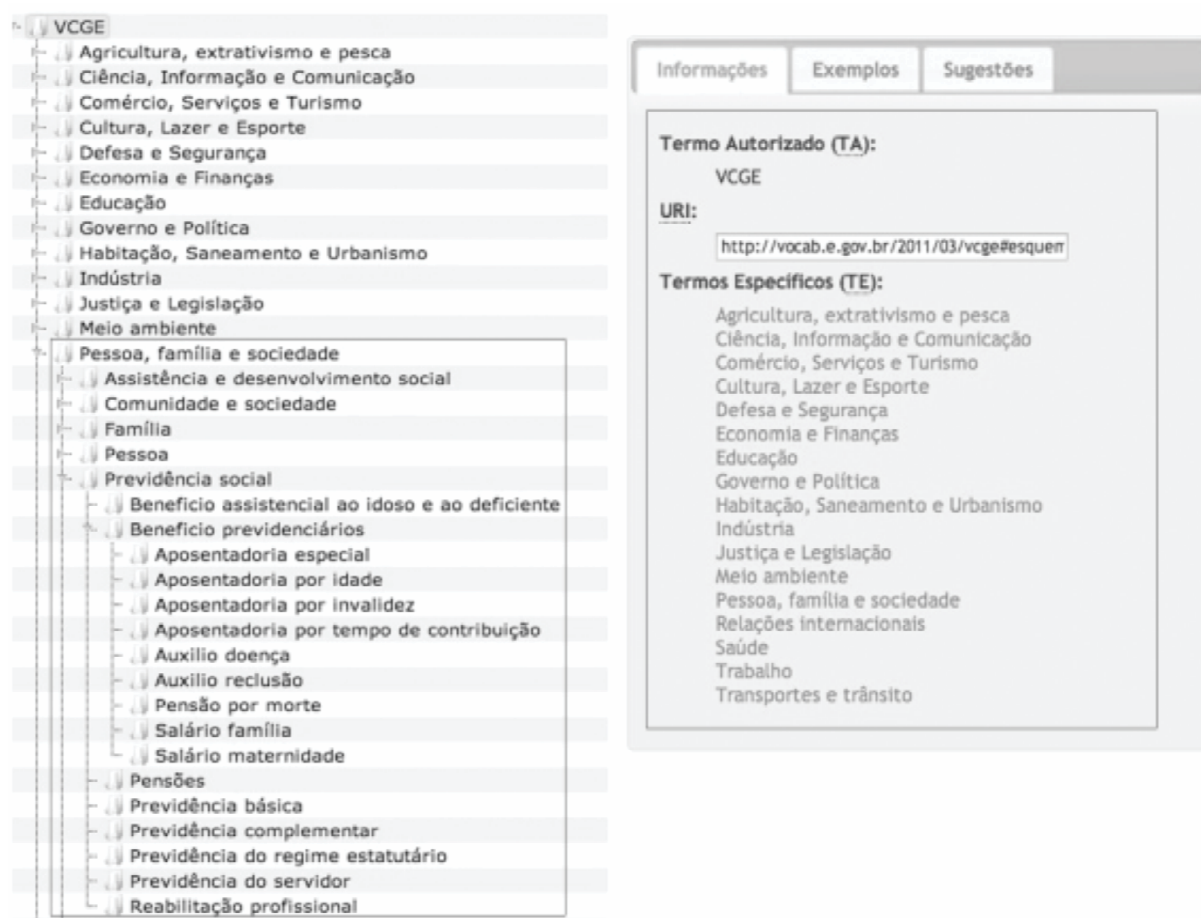


Figura 1. Previdência Social no Vocabulário Controlado do Governo Eletrônico.

Fonte: Brasil (2011).

Ainda sobre este ponto, foi explicitado que o foco na elaboração do Vocabulário foi facilitar seu uso pelos cidadãos (Brasil, 2011); assim, quanto mais evidente forem os termos, mais fácil será sua localização e utilização. Dar destaque a algumas áreas em detrimento de outras com o mesmo nível hierárquico, faz o Vocabulário perder a consistência entre as relações. Nesse sentido, este estudo propõe passar a classe de Previdência Social para um nível hierárquico superior, deixando-a no mesmo nível das outras classes acima mencionadas.

Ademais, foram encontradas inconsistências conceituais, referentes ao postulado da monorreferencialidade. Segundo tal concepção, a denominação de uma unidade de conhecimento - conceito (Dahlberg, 1978), "é um termo o qual guarda com ele uma relação unívoca, isto é, para cada conceito existe apenas uma

denominação e cada denominação vale apenas para um conceito" (Gomes, 1990, p.19).

Nessa direção, entendem-se como diferentes os termos "Previdência do regime estatutário" e "Previdência do servidor", quando na verdade eles possuem uma relação de equivalência. A Previdência do regime estatutário possui como finalidade "garantir os direitos e benefícios previdenciários dos servidores" (Maranhão, 2012, p.4), enquanto a Previdência do servidor é aquela "obrigatória aos servidores públicos federais, estaduais, municipais e do Distrito Federal" (Brasil, 2013a, *online*). Observa-se ainda que o termo "Previdência do regime estatutário" não é facilmente encontrado no domínio da Previdência Social, sendo mais recorrente o termo "Previdência do servidor".

A relação de equivalência não está expressa pelo Vocabulário analisado, apesar de assim estar descrito na

sua metodologia de elaboração. Logo, conforme observado, há dois termos diferentes para representar um único conceito, perdendo-se assim o controle linguístico dos termos, que é um dos princípios básicos de qualquer vocabulário controlado. Embora na linguagem natural não existam sinônimos perfeitos, deve-se buscar na linguagem especializada o controle de sinônimos ou quase-sinônimos (Gomes, 1990).

Conforme mencionado por Moreiro González (2011), uma das características das taxonomias é não incluir definições para os termos, somente as relações hierárquicas. O VCGE não possui definições para os termos, embora na aba de sugestões seja solicitada a colaboração dos usuários do vocabulário para ajudarem a compor as notas de escopo (NE) com definição dos termos. Além disso, as definições constam na metodologia como uma das etapas de elaboração do Vocabulário. Percebe-se, pois, o desejo de tornar o VCGE ainda mais específico e com relações mais bem definidas.

Assim, este estudo sugere que se acrescentem aos termos as notas de escopo, o que facilitará a compreensão de cada termo representado. Como afirma Gomes (1990, p.25), o registro de definições sobre elementos “fornece as características do conceito que vão permitir seu agrupamento e indicar as relações”. Portanto, é possível afirmar que as definições trarão segurança ao organizador do tesouro no momento de estabelecer as relações (Gomes, 1990).

As propostas nascidas da presente pesquisa visam contribuir para que o VCGE se constitua um instrumento dinâmico de organização do conhecimento, podendo ser utilizado para diversos fins, de acordo com os interesses de cada usuário. Com o desenvolvimento dos conceitos, da estrutura e das relações para os respectivos elementos, este instrumento poderá se tornar, no futuro, um dos pilares de sustentação para a elaboração de uma ontologia no âmbito do Governo brasileiro.

Discussão

Como discutido no referencial teórico deste trabalho, é necessário incorporar a prática de reutilização sistemática, de forma a facilitar a comunicação e sua recuperação pela sociedade. No caso aqui apresentado, foram

pesquisados vocabulários que representassem o domínio da Previdência Social brasileira, com o propósito de reutilizar suas estruturas e conceitos.

No entanto, observam-se no VCGE alguns pontos que tornaram o Vocabulário inconsistente, principalmente para a ligação semântica entre os dados publicados pela Dataprev.

Ademais, é preciso definir vocabulários que representem os dados publicados, de forma a garantir sua legibilidade por máquina e compartilhar as aplicações. Um vocabulário que se aplica à estrutura do VCGE é o *Simple Knowledge Organization System* (SKOS), que pode ser utilizado em conjunto com o *Resource Description Framework* (RDF) para representações semiformais de sistemas de organização do conhecimento, como tesouros, taxonomias, esquemas de classificação e listas de cabeçalhos de assunto. Fazendo uso de formalismo RDF, as representações se tornam legíveis por máquina e podem ser trocadas entre aplicativos de *software* e publicadas na *Web* (Putkey, 2011).

Analisar e propor uma revisão do vocabulário demandará tempo e investimento em trabalhos e pesquisas, além de depender, fundamentalmente, da participação de novos atores da Previdência Social, o que foge ao escopo deste estudo.

É preciso que haja apoio dos órgãos envolvidos para uso prático das tecnologias da Web Semântica, para tornar efetiva e reutilizável pelos cidadãos a publicação dos dados abertos do setor público brasileiro. A racionalização dos esforços, por meio de instrumentos já existentes de organização do conhecimento, pode impulsionar o desenvolvimento de novas pesquisas que viabilizem maior participação da sociedade no que tange à transparência e à *accountability* das informações. Portanto, são necessárias parcerias entre os órgãos governamentais, a iniciativa privadas e a sociedade, para a publicação e reutilização dos dados abertos, pois estes se tornam mais poderosos quanto mais usos se façam deles (Byrne & Goddard, 2010).

O Plano de Ação para Infraestrutura Nacional de Dados Abertos (INDA) prevê a prospecção de ontologias e vocabulários existentes a serem reutilizados para publicação de dados governamentais abertos. O *site* <dados.gov.br> funciona como um catálogo coletivo, re-

colhendo dados dos mais diversos órgãos públicos. Dessa forma, os “dados do catálogo serão classificados e agrupados conforme os temas (assuntos) constantes no Vocabulário Controlado de Governo Eletrônico (VCGE)” (Brasil, 2012b, p.24). Apesar de existir expectativa em se utilizar o VCGE para organização semântica dos dados publicados, este estudo demonstrou que o vocabulário em questão precisa ser revisto e alterado, a fim de aprimorar a organização conceitual das temáticas envolvidas.

Diante do exposto, é nesse contexto de utilização do VCGE como instrumento de organização do conhecimento utilizado na publicação dos dados abertos do Estado brasileiro, que se fundamenta a presente proposta de revisão da classe da Previdência Social. Utilizando-se a metodologia e os resultados aqui apresentados, faz-se necessário que outros órgãos públicos, além da Dataprev, interessem-se em analisar e propor alterações e atualizações, nas classes cuja temática contemple os dados por eles geridos.

Conclusão

Embora existam leis que garantem o acesso às informações públicas, estas são disseminadas de forma dispersa e com baixa utilização pela sociedade. Elas ainda são recursos pouco explorados, devido à dificuldade de busca, processamento e reutilização.

A população, que é produtora de dados, também é consumidora. Por essa razão, o Estado, e, aqui em espe-

cial, a Previdência Social, precisa aprimorar os serviços já existentes e ao mesmo tempo elaborar outros novos, que contribuam para uma relação de simbiose entre ela e a sociedade. A publicação de seus dados de forma aberta possibilitará a criação, por parte da própria sociedade, de novos serviços e informações que contribuirão para a construção da cidadania, mediante maior acesso às informações e, consequentemente, a construção de uma sociedade mais participativa e consciente.

Todavia, muitos desafios ainda precisam ser vencidos, como a ausência de vocabulários controlados e de modelos conceituais, a inexpressiva cooperação entre as áreas públicas para compartilhamento de experiências, a falta de semântica entre os dados e a ineficaz divulgação dos dados já publicados.

Contra esse quadro, a Dataprev tem o papel de apoiar a evolução dos assuntos pertinentes à Previdência Social, por meio das tecnologias de informação. Tanto a metodologia apresentada quanto os resultados e a discussão devem instrumentalizar e apoiar os projetos em curso na Dataprev, especialmente as ações ligadas ao trabalho do Governo Eletrônico (Projetos Padrões de Interoperabilidade de Governo Eletrônico (*e-Ping*) e INDA).

Diante do exposto, espera-se, em trabalhos futuros, conseguir mostrar a aplicação das propostas feitas ao VCGE. Espera-se, também, que outras classes além da Previdência Social sejam objeto de estudos neste sentido. O Vocabulário pode ser reutilizado para diversos fins no âmbito governamental, sendo importante a realização de pesquisas na área, com reflexos para a sociedade em geral.

Referências

Aganette, E.C.; Alvarenga, L.; Souza, R.R. Taxonomias corporativas: um estudo sobre definições e etapas de construção fundamentado na literatura publicada. In: Encontro Nacional de Pesquisa em Ciência da Informação, 11., 2010, Salvador. *Anais...* Brasília: ANCIB, 2010. p.1-28.

Berners-Lee, T. *Linked data*. 2006. Available from: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Cited: Jul. 11, 2012.

Berners-Lee, T. *Putting government data online*. 2009, p.1-7. Available from: <<http://www.w3.org/DesignIssues/GovData.html>>. Cited: Jul. 11, 2012.

Berners-Lee, T.; Hendler, J.; Lassila, O. The semantic web. *Scientific American*. 2001, p.29-37. Available from: <[\[sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21\]\(http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21\)>. Cited: Dec. 15, 2012.](http://www.</p>
</div>
<div data-bbox=)

Brasil. Ministério do Planejamento, Orçamento e Gestão. *Vocabulário controlado do governo eletrônico*. 2011. Disponível em: <<http://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padrees-de-interoperabilidade/vcge>>. Acesso em: 1 jan. 2012.

Brasil. Ministério da Previdência Social. *Anuário estatístico da previdência social*: AEPS 2011. v.20, 2012a. Disponível em: <<http://www.previdencia.gov.br/estatisticas/aeps-2011-anuario-estatistico-da-previdencia-social-2011-2/>>. Acesso em: 2 maio 2013.

- Brasil. Ministério do Planejamento, Orçamento e Gestão. *Plano de ação da infraestrutura nacional de dados abertos*. Brasília: Secretaria de Logística e Tecnologia da Informação, 2012b.
- Brasil. Ministério do Planejamento, Orçamento e Gestão. *Serviços*. 2013a. Disponível em: <<http://www.brasil.gov.br/para/servicos/aposentadoria/previdencia-do-servidor>>. Acesso em: 22 jul. 2013.
- Brasil. Controladoria Geral da União. *Estrutura organizacional do Governo Federal*. 2013b. Disponível em: <<http://www.cgu.gov.br/publicacoes/prestacaocontaspresidente/2010/Arquivos/1.01.pdf>>. Acesso em: 22 jul. 2013.
- Byrne, G.; Goddard, L. The strongest link: Libraries and linked data. *D-Lib Magazine*, v.16, n.11/12, 2010. Available from: <<http://www.dlib.org/dlib/november10/byrne/11byrne.html>>. Cited: Jul. 29, 2012.
- Campos, M.L.A.; Gomes, H.E. Taxonomia e classificação: o princípio de categorização. *DataGramaZero: Revista de Ciência da Informação*, v.9, n.4, 2008. Disponível em: <http://www.dgz.org.br/ago08/F_I_art.htm>. Acesso em: 13 jan. 2013.
- Chahin, A. et al. *E-Gov.br: a próxima revolução brasileira: eficiência, qualidade e democracia: o governo eletrônico no Brasil e no mundo*. São Paulo: Prentice Hall, 2004.
- Dahlberg, I. A referent-oriented analytical concept theory of interconcept. *International Classification*, v.5, n.3, p.142-150, 1978.
- Gomes, H.E. (Coord.). *Manual de elaboração de tesouros monolíngües*. Brasília: Programa Nacional de Bibliotecas das Instituições de Ensino Superior, 1990.
- Maranhão (Estado). Ministério Público. Procuradoria Geral de Justiça. *Plano plurianual 2012-2015*. 2012. Disponível em: <http://www.mp.ma.gov.br/arquivos/arquivos_site_antigo/Noticia6885_A6285.pdf>. Acesso em: 22 jul. 2013.
- Marcondes, C.H.; Campos, M.L.A. Ontologia e Web Semântica: o espaço da pesquisa em Ciência da Informação. *PontodeAcesso*, v.2, n.1, p.107-136, 2008. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/viewArticle/2669>>. Acesso em: 28 jun. 2011.
- McGuinness, D.L. Ontologies come of age. In: Fensel, D. (Ed.). *Spinning the semantic web: Bringing the world wide web to its full Potential*. Massachusetts: MIT Press, 2003. Available from: <[http://www-ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-\(with-citation\).htm](http://www-ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm)>. Cited: Jan. 21, 2013.
- Méndez, E.; Greenberg, J. Linked data for open vocabularies and HIVE's global framework. *El Profesional de la Información*, v.21, n.3, p.236-244, 2012. Available from: <<http://dx.doi.org/10.3145/epi.2012.may.03>>. Cited: Mar. 25, 2013.
- Moreiro González, J.A. *Linguagens documentárias e vocabulários semânticos para web: elementos conceituais*. Salvador: EdUFBA, 2011.
- Neves Júnior, O. *Identificação e descrição de fatores críticos de sucesso em iniciativas brasileiras de telecentros*. 2006. Dissertação (Mestrado em Gestão do Conhecimento e Tecnologia da Informação) - Universidade Católica de Brasília, Brasília, 2006. Disponível em: <http://www.btdt.ucb.br/tede/tde_arquivos/3/TDE-2007-03-06T120657Z-402/Publico/Odilon%Neves%20Jr.pdf>. Acesso em: 6 nov. 2013.
- Open Government Data: 8 principles of open government data. *Public.Resource.org*. Sebastopol, 2007. Available from: <<http://www.opengovdata.org/home/8principles>>. Cited: Jan. 15, 2013.
- Putkey, T. Using SKOS to express faceted classification on the semantic web. *Library Philosophy and Practice*, Lincon, Jan. 2011. Available from: <<http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1652&context=libphilprac>>. Cited: Mar. 1, 2013.
- Ribeiro, C.J.S.; Almeida, R.F. Dados abertos governamentais (Open Government Data): instrumento para exercício de cidadania pela sociedade. In: Encontro Nacional de Pesquisa em Ciência da Informação, 12., 2011, Brasília. *Anais...* Brasília: ANCI, 2011. p.2568-2580.
- Souza, R.R.; Alvarenga, L. A web semântica e suas contribuições para a ciência da informação. *Ciência da Informação*, v.33, n.1, p.132-141, 2004. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/50/50>>. Acesso em: 13 jan. 2013.
- Vaz, J.C.; Ribeiro, M.M.; Matheus, R. Dados governamentais abertos e seus impactos sobre os conceitos e práticas de transparência no Brasil. *Cadernos PPG-AU/UFBA*, v.9, (Esp.), p.45-62, 2010. Disponível em: <<http://www.portalseer.ufba.br/index.php/ppgau/article/view/5111>>. Acesso em: 8 jan. 2013.
- Vital, L.P.; Café, L.M.A. Ontologias e taxonomias: diferenças. *Perspectivas em Ciência da Informação*, v.16, n.2, p.115-130, 2011. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/200/927>>. Acesso em: 21 jan. 2013.
- World Web Consortium Brasil. *Manual dos dados abertos: governo*. 2011. Disponível em: <http://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf>. Acesso em: 4 jan. 2013.
- World Web Consortium Brasil. *Melhorando o acesso ao governo com melhor uso da Web*. São Paulo: W3C, 2009. Disponível em: <<http://www.w3c.br/divulgacao/pdf/gov-web.pdf>>. Acesso em: 8 jan. 2013.