

Comunicação Oral

EM BUSCA DE UMA INTEROPERABILIDADE SEMÂNTICA NA PUBLICAÇÃO DE DADOS ABERTOS GOVERNAMENTAIS: A EXPERIÊNCIA SOBRE OS DADOS DA PREVIDÊNCIA SOCIAL

Claudio Jose Silva Ribeiro – UNIRIO
Durval Viera Pereira – UNIRIO

Resumo

Objetiva discutir à luz da filosofia analítica a classificação dos elementos no domínio da Previdência Social, propondo ainda um conjunto de ações para representação destes elementos, com vista a contribuir para a interoperabilidade semântica na publicação de dados abertos governamentais. Constitui-se em uma pesquisa aplicada, desenvolvida à luz da Filosofia Analítica para o entendimento e estruturação do conhecimento. Como marco empírico para delimitação e análise dos elementos do domínio foram utilizadas as informações do Anuário Estatístico de Acidente do Trabalho e do Anuário Estatístico da Previdência Social. Os resultados observados contemplaram: a análise da sintaxe e semântica dos termos do domínio, a organização do conjunto de classes segundo uma abordagem dedutiva e a questão da interoperabilidade semântica, pois esta necessita de um vocabulário que forneça significado aos dados publicados. Consta que a Filosofia Analítica se demonstrou adequada como teoria para o entendimento e estruturação do conhecimento relativo aos domínios concernentes às bases de dados, contribuindo para o avanço rumo à efetiva interoperabilidade semântica entre dados provenientes de bases governamentais heterogêneas.

Palavras-chave: Filosofia Analítica. Interoperabilidade semântica. Dados governamentais Abertos. Previdência Social.

Abstract

This report discusses principles of analytic philosophy in the classification of the elements in the field of Social Security from Brazil, even proposing a set of actions for the representation of these elements in order to contribute to the semantic interoperability in the publication of open government data. It constitutes an applied research, using the principles of Analytic Philosophy for understanding and structuring of knowledge. In the empirical field was used the Anuário Estatístico de Acidente do Trabalho and the Anuário Estatístico da Previdência Social for delimitation and analysis of elements of the domain. The results observed included: analysis of the syntax and semantics of the terms of the domain, the organization of terms classes according to deductive approach and addressing to the problem of semantic interoperability, as requires a vocabulary that gives meaning to the data published. Indicates that Analytic Philosophy is adequate as a theory for understanding and structuring of knowledge relating to the fields pertaining to databases, to improve semantic interoperability between data from heterogeneous databases government.

Keywords: Analytic Philosophy. Semantic interoperability. Open government data. Social Security.

1 INTRODUÇÃO

As linguagens e teorias relativas ao signo e ao significado têm sido utilizadas para fornecer o embasamento teórico-metodológico, de forma a auxiliar na representação dos elementos do mundo real (APEL, 2000). Uma das áreas que se debruça sobre este tema é a Ciência da Informação e, em especial, a Organização da Informação e do Conhecimento, que busca fornecer modelos para estudo de usuários e instituições, fazendo uso de serviços de informação em diferentes comunidades discursivas.

Nesta direção, Hjørland (2003) observa que a visão da teoria da semiótica social é que significado, signos e documentos são desenvolvidos para funcionar na relação para padronizar práticas informacionais em comunidades. Isto nos lembra Frohmann (2012), quando este explica que é fundamental conhecer as questões históricas que envolvem as forças sociais e políticas que configuram as práticas dos documentos analisados. Assim, aparece o conceito de informatividade dos documentos, que é reconhecida como algo dependente de práticas, mas também é dependente do que lhes dá forma e os configuram (FROHMANN, 2012).

Por outro lado, para desenvolver projetos de aplicações computadorizadas, Sowa (2000) observa a importância de representar o conhecimento no nível linguístico, compreendendo conceitos, palavras e expressões em linguagem natural.

Dentre as abordagens epistemológicas que trabalham com conceitos e significados, uma corrente que se destaca é a Filosofia Analítica, caracterizada “pela concepção de que a lógica e a teoria do significado ocupam um papel central na filosofia, sendo que a tarefa básica é a análise lógica das sentenças, através da qual se obtém a solução dos problemas filosóficos” (JAPIASSÚ; MARCONDES, 2006).

Assim, apoiados na pesquisa de Ribeiro (2008), partimos do pressuposto que a Filosofia Analítica pode auxiliar na busca por uma interoperabilidade semântica, pois a análise lógica das sentenças alvo de representação pode viabilizar a convergência de conceitos e significados entre domínios distintos. Cabe destacar que esta interoperabilidade é entendida como a capacidade de diferentes sistemas de informação trocarem informações de acordo com o significado da informação codificada (como pretendido pelos criadores e mantenedores do sistema de informação) (PATEL et al., 2004).

Adicionalmente, o esforço de publicação de dados governamentais abertos¹ na Previdência Social aponta para a "organização, revisão e disponibilização de todo o vocabulário relacionado ao conteúdo a ser publicado" (RIBEIRO; ALMEIDA, 2011). Ademais, a estruturação semântica aplicada nestes dados aliada ao uso de vocabulários e formalismos para representação dos elementos sob análise, pode tornar viável o processamento automatizado das informações e conteúdo, pois este é um dos caminhos naturais para trabalharmos com a miríade de dados que estão presentes atualmente na Internet (RIBEIRO; ALMEIDA, 2011).

Organizar os dados governamentais, facilitar sua recuperação e sua utilização, são objetivos da abertura dos dados governamentais, que vão ao encontro de uma das visões sobre o objeto da Ciência da Informação. Esta visão é percebida como "ajudar as pessoas (ou mais amplamente: o atores) que estão confusas com a situação do uso do conhecimento (e que ficarão mais confusas por causa do modelo de sociedade pós-moderna) (WERSIG, 1993).

O presente relato é continuidade da pesquisa iniciada em 2011, na qual foram esclarecidos os benefícios dos dados governamentais abertos para o exercício da cidadania pela sociedade, além de apresentar o projeto desenvolvido pela Empresa de Tecnologia e Informações da Previdência Social (Dataprev) na publicação de dados sobre acidentes de trabalho (RIBEIRO; ALMEIDA, 2011). Entretanto, naquela época, os estudos sobre a semântica aplicada aos dados da Previdência ainda estavam em nível inicial, carecendo de instrumentos para a representação dos dados e informações publicadas. Desde então, alguns métodos e teorias têm sido experimentados, na tentativa de solucionar esse problema de falta de semântica dos dados.

Desta forma, o objetivo deste trabalho é discutir à luz da Filosofia Analítica a estruturação de elementos no domínio sob análise e propondo ainda um conjunto de ações para representação destes elementos. Esta representação contribuirá para utilizá-los [os elementos do domínio] em projetos futuros, de forma a viabilizar o processo de interoperabilidade semântica entre ambientes na Web.

¹ W3C (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".

2 A NECESSIDADE DE INCORPORAR INTEROPERABILIDADE A SEMÂNTICA NO CONJUNTO DE DADOS PUBLICADOS

Um dos fatores motivadores para o tema da interoperabilidade foi a publicação da Carta de Serviços ao Cidadão. Os esforços do Governo em reduzir a exigência de documentos já constantes em bases de dados oficiais, impulsionaram o conjunto de ações na direção de buscar interoperabilidade entre as bases oficiais do Governo Brasileiro (RIBEIRO; ALMEIDA, 2011).

Por outro lado, é lícito supor que a *Web* alterou radicalmente a forma de compartilhar conhecimentos, pois um dos maiores fatores de sucesso deste ambiente é a liberdade que seus usuários têm em criar e disseminar informação a qualquer hora e em qualquer lugar. Ligações entre documentos através de estruturas de hipertexto permitem aos usuários percorrer o espaço informacional onde os documentos estão publicados, por meio dos assim chamados navegadores *Web*, enquanto motores de busca associados com analisadores de ligações de documentos fazem com que os usuários localizem os documentos requeridos. Esta combinação foi a principal responsável pelo sucesso da Internet e pelo seu pleno constante incremento. (BERNES-LEE, 2009).

Com a adoção de novas aplicações, como as redes sociais associadas a plataformas tecnológicas mais avançadas, novas culturas vão sendo estabelecidas sobre a rede mundial de computadores, possibilitando o compartilhamento de informações e dados, permitindo ainda a construção de ações colaborativas.

Diante deste contexto, Fonseca, Egenhofer e Borges (2000) observam que a interoperabilidade semântica pode ser capaz de solucionar alguns problemas relacionados à complexidade de determinados domínios ou áreas de representação. Nestes domínios é possível nos depararmos com fatos que poderão ter diferentes descrições. Com ações ligadas à interoperabilidade semântica será possível proporcionar um uso mais efetivo das informações disponíveis na *Web*, pois tornará viável a reconciliação das definições estabelecidas para os respectivos fatos identificados.

Esta reconciliação semântica traz "consistência e integridade aos recursos de informação e serviços de dados" (RIBEIRO, 2008) uma vez que está associado à compatibilidade conceitual, que "é um mapeamento da potencialidade semântica das linguagens estudadas, fornecendo os resultados da análise de compatibilidade entre linguagens sob os pontos de vistas semântico e estrutural" (CAMPOS, 2005).

Já na visão de Lima (2010), os possíveis problemas de interoperabilidade podem ser categorizados em quatro tipos de heterogeneidade: (1) Sistema – incompatibilidades entre

hardware e sistemas operacionais; (2) Sintática – diferenças na codificação e representação da informação; (3) Estrutural – variações nos modelos de dados e esquemas; (4) Semântica – inconsistências na terminologia e significados.

No âmbito desta pesquisa, trataremos os elementos que ajudam a viabilizar as formas de interoperabilidade: sintática e semântica. Vale esclarecer que a interoperabilidade sintática se baseia na “codificação dos dados, mediante utilização das linguagens de marcação para desenvolvimento de sistemas, modelos de gestão de documentos e registros eletrônicos, e formatos de apresentação da informação” (MUCHERONI; SILVA, 2011). Já a interoperabilidade semântica se refere a “habilidade de dois sistemas se comunicarem, tendo entre eles a mesma interpretação e, portanto, as mesmas implicações lógicas sobre as informações trocadas” (MIRANDA, 2011).

Adicionalmente, em busca de construir instrumentos que viabilizem a plena implementação da interoperabilidade entre aplicações do Governo Brasileiro, também foram desenvolvidos alguns marcos normativos, que orientam as ações das unidades operacionais encarregadas de implementar as políticas de Governo. A constatação da necessidade destes marcos está explícita no relato Panorama da Interoperabilidade no Brasil, publicado pelo Ministério do Planejamento e destacado a seguir:

[...] para implantação efetiva de determinado padrão tecnológico, tem-se prontamente o conjunto de especificações necessárias e condições de uso construídas e públicas. No entanto, a implantação efetiva de um serviço interoperável não estará pronta pela simples definição do XML *Schema*. Além de catalogar o serviço, será necessário equacionar toda uma gama de condições institucionais: seja no âmbito da disponibilização da infraestrutura e da definição de custos de acesso ao serviço, seja na harmonização da semântica entre as diferentes agências envolvidas ou ainda em ajustamentos de processos que se fizerem necessários. Em síntese, em geral, a adoção de um padrão tecnológico envolve menor número de atores e temas, enquanto a mudança de abordagem na prestação de serviços de governo eletrônico com vistas a agregar os benefícios *in totum* da interoperabilidade exigirá acordo com grupos de gestores, muitas vezes distantes do processo de construção do marco normativo (BRETAS, 2010).

As ações decorrentes do movimento em busca da interoperabilidade, também impulsionaram a criação da Infraestrutura Nacional de Dados Abertos - INDA (RIBEIRO; ALMEIDA, 2011). Cabe lembrar que a INDA é um esforço que visa promover o desenvolvimento do conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas no modelo de Dados Abertos. Composta por quatro Grupos de

Trabalho (GT's) que objetivam: a criação de normas e padrões; avaliação de ações em curso; uso de formalismos para representação; e convergência tecnológica. Estes GT's vêm trabalhando na pavimentação do caminho que reúne os aspectos de interoperabilidade em conjunto com as demandas para uso de dados abertos (GRUPO DE TRABALHO DA INFRAESTRUTURA NACIONAL DE DADOS ABERTOS, 2013).

Complementarmente, o Governo Brasileiro estabeleceu políticas gerais da e-PING específicas sobre a interoperabilidade semântica. Estas políticas são divididas em três pontos principais (BRASIL, 2013):

- **Desenvolvimento e manutenção de ontologias e outros recursos de organização da informação:** visando facilitar o cruzamento de dados de diferentes fontes de informação, quando da sua utilização por outras organizações integrantes da administração pública, por organizações da sociedade civil ou pelo cidadão, devem ser utilizados recursos tais como vocabulários controlados, taxonomias, ontologias e outros métodos de organização e recuperação de informações.
 - **Desenvolvimento e adoção de um padrão de modelagem de dados para Governo:** baseada em notação simples, objetiva e facilmente utilizável, a modelagem deve: evidenciar as integrações atuais e as integrações necessárias entre os dados; apoiar as interações do Governo em suas diversas secretarias e órgãos; apoiar o alinhamento com os processos de negócios governamentais; promover a melhoria na gestão pública; e servir como arquitetura de interoperabilidade para o Governo.
 - **Desenvolvimento e adoção de uma política de disseminação de dados e informações:** baseada em experiências internacionais de abertura de dados governamentais (*OpenData*), a política consiste em uma série de ações coordenadas para orientar a incorporação de processos de disponibilização dos dados públicos para permitir seu melhor uso pela sociedade, alinhada com a diretriz da e-PING de adoção de padrões abertos na interação do Governo Federal com a sociedade.
- Com nossa pesquisa centrada no ambiente Web, podemos inferir que para seguir estas políticas será necessário dotar a *Web* de facilidades para permitir o processamento automatizado das informações e conteúdo. Percebemos também, que este é um dos caminhos naturais para trabalharmos com a miríade de dados que estão presentes atualmente na Internet e dentro deste contexto que se insere este relato.

3 PROCEDIMENTOS METODOLÓGICOS

Podemos classificar o presente relato como uma pesquisa aplicada, pois tem como objetivo reunir e elaborar informações para “contribuir para fins práticos, visando à solução mais ou menos imediata do problema encontrado na realidade” (BARROS; LEHFELD, 2000).

Desenvolvemos, à luz da filosofia analítica, buscar entender a estruturação e relacionamentos do conhecimento na área de acidente de trabalho dentro do campo da Previdência Social brasileira.

Como marco empírico para delimitação e análise dos elementos do domínio foram utilizados o AEAT – Anuário Estatístico de Acidente do Trabalho – (ANUÁRIO..., 2010) e o AEPS – Anuário Estatístico da Previdência Social – (ANUÁRIO..., 2011). O AEAT faz parte do AEPS e contém uma análise dos registros de Acidente de Trabalho segundo um viés de agregações e tabulações estatísticas. Estes registros não trazem informações individualizadas, pois o objetivo da publicação é auxiliar a formulação de políticas previdenciárias para tratar os Benefícios decorrentes dos eventos de Acidente de Trabalho. Tanto o AEAT quanto o AEPS já estão disponíveis na internet com ferramentas que permitem a extração de dados, mas estas ainda não estão dotadas de facilidades para descrição semântica (RIBEIRO; ALMEIDA, 2011).

4 O USO DA FILOSOFIA ANALÍTICA PARA AUXÍLIO À CAPTURA DA SEMÂNTICA

Como ponto de partida, decidiu-se que, no âmbito desta pesquisa, era necessário buscar apoio no entendimento das “coisas do mundo”. Ao longo do tempo, este entendimento passou por diversas transformações e teve diferentes abordagens epistemológicas para a realização da pesquisa. Desde Platão até Husserl, os objetos da investigação filosófica tradicional sempre estiveram ligados ao conhecimento, ao ser, à existência, à verdade, à liberdade, ao bem e ao belo. A principal preocupação de observação e entendimento, sempre esteve relacionada ao caráter mais geral e abstrato do que simplesmente com questões concretas e particulares, tais como os objetos que nos circundam em nosso cotidiano (COSTA, 1992).

Por outro lado, a interação entre fonte e receptor, além da representação dos dados e informações que serão alvo de disseminação e comunicação, indicam o caminho para o início da jornada na direção do entendimento e da semântica. Desde a observação dos estudos da metafísica (onde a visão era marcada pelo objetivismo), até mais recentemente, à filosofia contemporânea, marcada pela virada linguística (na qual a linguagem passa a ser o novo paradigma de pesquisa em filosofia), foi possível perceber não apenas o deslocamento do foco dos estudos na direção da linguagem (CHAUI, 1995), mas também o conjunto de questões ligadas à semiótica e o problema do significado (ERIKSSON, 1999). Nesta direção, as visões trazidas e apoiadas na Filosofia Analítica (APEL, 2000; COSTA, 1992), fizeram com que a

exploração dos fundamentos da filosofia e do entendimento do processo informacional, nos auxiliassem a demonstrar que para representar melhor a informação, tínhamos que passar pela análise dos aspectos de comunicação entre os envolvidos no ato de comunicar (BERNOIT, 2002).

O entendimento do conhecimento presente no universo de informação passa pela percepção da estruturação deste contexto, que é organizado em uma rede de nós e ligações. Cada um destes elementos deve possuir uma especificação, composta por nome, descrição de objetivos e metas. Os nomes devem estar descritos por sentenças curtas, que caracterizam o atendimento a estas metas. A descrição dos objetivos está associada a uma especificação teórica, que compreende um conjunto de expressões formuladas em linguagem que represente o detalhamento para atendimento às metas (RIBEIRO, 2008).

Portanto, a complementação da especificação das expressões foi desenvolvida à luz da Filosofia Analítica, onde a relação entre o entendimento da realidade e a linguagem utilizada, esteve mediada pela sintaxe, semântica e pragmática, por meio de (APEL, 2000):

- Sintaxe, observando a relação intralinguística dos signos utilizados;
- Semântica, ressaltando a relação dos signos com os fatos extralinguísticos designados;
- e
- Pragmática, notando a relação dos signos com os usuários da linguagem, objetivando seu uso concreto.

Nestas três dimensões percebe-se uma síntese dos motivos centrais da filosofia linguística-analítica ocidental contemporânea, na medida em que esta se aproxima de uma teoria geral dos signos proposta por Morris (JAPIASSU; MARCONDES, 2006).

Assim, mediado pela linguagem, o processo de representação ganha destaque na medida em que objetiva designar corretamente as qualidades conhecidas à luz de seus significados, que são percebidos, apreendidos e entendidos na análise dos conteúdos do domínio (APEL, 2000).

5 ANÁLISE DOS RESULTADOS

Com a compreensão e delimitação do domínio sob análise, verificamos os resultados descritos a seguir.

5.1 O PROCESSO DE ANÁLISE E REPRESENTAÇÃO DOS ELEMENTOS

Para publicar informações na *Web* é preciso realizar o entendimento e representação de informação. Contudo, sabemos que entender e representar são apenas partes do problema,

pois estas características, além de permitirem a melhor disseminação do conhecimento, auxiliam no processo de recuperação de documentos, ou qualquer outro recurso, que estejam disponíveis.

A análise da semântica dos termos do domínio permitiu que, além da montagem do dicionário terminológico, houvesse a organização do conjunto de classes segundo uma abordagem dedutiva, com a estruturação da hierarquia de conceitos e suas associações. Relações de especialização foram identificadas e registradas. A representação a seguir demonstra a hierarquia obtida:

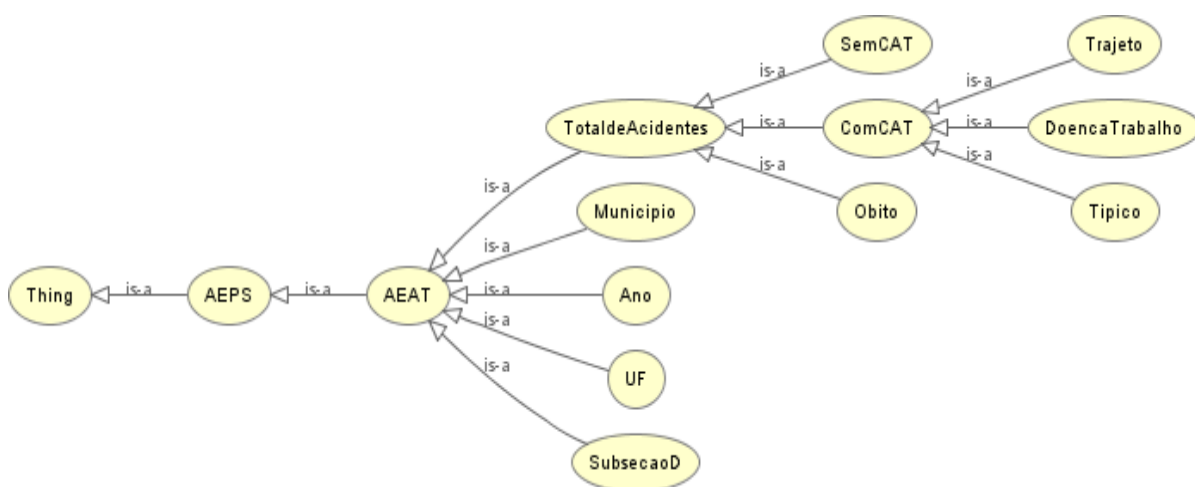


Figura 1: Estrutura de classes para os Anuários (Previdência e Acidentes de Trabalho)
Fonte: O autor.

Por outro lado, as especificações das classes e a posterior identificação e descrição dos atributos ou características destas classes, se apoiam na definição sintática das especificações, bem como na visão da relação com os usuários que possibilitaram a validação destas especificações. Posteriormente, por meio de uma abordagem indutiva, foram reunidos os elementos comuns, utilizando critérios de similaridade entre as especificações obtidas. As definições das classes serviram para corroborar a estrutura obtida anteriormente, conforme apresentado na Figura 1.

5.2 INTEROPERABILIDADE SEMÂNTICA E A NECESSIDADE DE VOCABULÁRIOS

Para Marcondes e Sayão (2008), diferentes conteúdos disponibilizados na Internet “só poderão ser integrados e reusados, no sentido de terem aproveitadas as sinergias uns dos

outros, se estiverem ancorados por sistemas que permitam um alto grau de interoperabilidade”.

Assim, é lícito supor que para o processo de interoperabilidade semântica, é necessário utilizar instrumentos para representação da informação armazenada nos dados e sistemas. Neste processo de representação devemos destacar: os vocabulários controlados, as taxonomias, os padrões de metadados e as ontologias. Desta forma, podemos dizer que a interoperabilidade semântica depende da utilização de vocabulários bem estruturados e fundamentados, estabelecendo relações semânticas entre os termos e proporcionando que máquinas consigam realizar inferências sobre os dados representados. 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 de 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.). Esta abordagem permite que os processos de metadados familiares, enquanto a capacidade de vincular os dois vocabulários e registros de dados si fornece uma infra-estrutura que permite a descoberta de informação mais eficaz e uso. (MÉNDEZ, GREENBERG, 2012, p. 237, tradução nossa)

No tema de Dados Abertos, a INDA recomenda a utilização do Vocabulário Controlado do Governo Eletrônico (VCGE) para atribuição de semântica aos dados governamentais publicados no Portal Dados.gov.br. (BRASIL, 2012). No entanto, o VCGE não possui relações semânticas entre as classes, além da relação gênero/espécie. Ademais, há alguns indícios de que o vocabulário em questão não consegue representar toda a complexidade presente no domínio coberto pelo Governo Brasileiro, o que pode vir a comprometer o trabalho de representação dos dados publicados.

Esta afirmação pode ser comprovada ao observar o campo empírico de estudo deste relato. Na classe referente à Previdência Social, representada no VCGE, observamos que a descrição temática feita não corresponde a completude dos termos e dos conceitos tratados pelo domínio em questão. Termos importantes para entendimento da Previdência Social não estavam contemplados no vocabulário, como por exemplo, os tipos de pensões concedidas.

Se um cidadão precisar buscar por uma espécie de pensão ao qual se beneficia, não encontrará representada esta pensão a qual procura, pois o termo “Pensão” não é classificado pelo VCGE. Atualmente, o Ministério da Previdência Social (ANUÁRIO, 2012) oferece as seguintes pensões: Pensão especial à 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; e Pensão por morte previdenciária.

Anuário Estatístico de Acidentes de Trabalho - AEAT

Ver Recursos (24) Relacionado (0) Histórico

Anuário Estatístico de Acidentes de Trabalho - AEAT, documento público, com edição e publicação anuais em meio papel e versão digital acessível por meio do site do Ministério da Previdência Social.

No caso deste trabalho, selecionamos informações disponibilizadas na Seção I subseção D do Anuário, conteúdo que trata de Comunicação de Acidente de Trabalho (CAT), suas consequências, por localização geográfica das ocorrências (município) e seus respectivos quantitativos.

As informações da Comunicação de Acidente do Trabalho (CAT) têm origem nos registros dos benefícios de natureza acidentária concedidos pelo INSS.

Para maiores detalhes de acesso (API), consulte a URL:
<http://api.dataprev.gov.br/previdencia/anuario/ANO/acidentes-do-trabalho.FORMATO>

Dois parâmetros são necessários para solicitar qualquer dado através da API: ANO e FORMATO. O parâmetro ano permite selecionar a série anual desejada, de 2002 a 2009. O formato define sua preferência em receber a resposta em XML, JSON ou CSV.

Campo	Valor
Fonte	http://api.dataprev.gov.br/doc/dadosDisp.htm
Autor	Autor não fornecido
Mantenedor	Mantenedor não fornecido
Assuntos	Segurança no Trabalho
Granularidade geográfica	municipal
Granularidade temporal	Anual
VCGE	Segurança no Trabalho [http://vocab.e.gov.br/2011/03/vcge#seguranca-trabalho]
Órgão - Esfera	Federal
Órgão - Poder	Executivo

Figura 2: Descrição temática dos dados de acidente de trabalho publicados no Portal Dados.gov.br.
 Fonte: Dados.gov.br (2013)

Cabe destacar ainda que a Dataprev iniciou o processo de publicação de dados abertos utilizando os dados sobre acidentes de trabalho, porém o termo “Acidente de trabalho” não é encontrado no VCGE. Os dados atualmente publicados no portal Dados.gov.br pela Dataprev, estão indexada com o termo “Segurança do trabalho”, como pôde ser observado na figura 2. Este termo dentro do VCGE é subordinado ao termo “Legislações trabalhistas”, que por sua vez é termo específico do termo “Trabalho”. Entretanto, os dados publicados sobre acidente de trabalho possuem uma temática específica, não devendo sua representação ser reduzida a legislações trabalhistas.

Vislumbramos 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 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.

6 DISCUSSÃO

Atendendo aos anseios da sociedade em geral, o esforço de publicação de dados abertos torna evidente a necessidade de prover informação para a melhor participação da sociedade na gestão do nosso país. No entanto, o maior desafio é dispor o alto volume de informação de maneira que seja possível a (re)utilização de todo este conjunto de dados, administrando da forma mais efetiva possível os conflitos de heterogeneidade (como por exemplo: iguais nomes que denotam diferentes conceitos (homônimos), diferentes nomes para mesmos conceitos (sinônimos, etc) entre as bases que hospedam os mesmos. Uma das maiores questões que surgem com este trabalho é como implementar um processo ágil de publicação que permita, inclusive, agregar semântica aos dados, e assim, conduza de uma *Web Sintática* para uma *Web Semântica*.

Nesta direção é possível destacar três questões que merecem especial atenção na continuidade desta pesquisa:

A primeira endereça para a abordagem epistemológica utilizada, pois o desenvolvimento e detalhamento do entendimento das expressões, classes e atributos, foi realizada à luz da filosofia analítica, por meio do tratamento da sintaxe, semântica e pragmática. No entanto, a exploração de outras vertentes do estudo filosófico-epistemológico, tal como a fenomenologia, pode ser alternativa viável para o processo de investigação da semântica, pois nessa abordagem a visão da essência dos problemas que são alvo de análise passa pelo uso de reduções fenomenológicas, onde as visões das essências imediatas são fixadas em dados, seja de maneira conceitual ou terminológica (HUSSERL, 2006).

A segunda questão endereça para o uso de vocabulários para a representação dos elementos, pois precisamos incorporar a prática de reutilização sistemática, de forma a facilitar a comunicação e a recuperação de informação pela sociedade. No caso aqui apresentado, buscamos vocabulários que representassem o domínio da Previdência Social brasileira, com o propósito de reutilizarmos suas estruturas e conceitos para publicação dos dados sobre as informações do Ministério da Previdência Social, salvaguardados pela Dataprev. No entanto, ao nos depararmos com o VCGE, percebemos alguns pontos que tornaram o vocabulário inconsistente, principalmente para o uso ao qual pretendíamos, ou seja, na ligação semântica entre os dados publicados pela Dataprev.

A terceira questão endereça para o modelo tecnológico a ser utilizado. Este modelo deve permitir que as bases de dados se integrem, além de possibilitar que haja interoperabilidade semântica entre os dados para que as informações demandadas sejam publicadas. Dentre os modelos tecnológicos a serem considerados, aquele que tem se

mostrado mais próximo aos esforços encaminhados é o assim chamado *hybrid ontology approach*, no qual ontologias locais referentes a cada base de dados são construídas a partir de um vocabulário controlado compartilhado (*shared vocabulary*). Desta forma, o mapeamento é feito e a interoperabilidade entra as bases efetivadas. (WACHE et al., 2001).

7 CONSIDERAÇÕES FINAIS

Apesar da aplicação da Filosofia Analítica ter se mostrado adequada à análise de um domínio específico, ainda existe um caminho a ser percorrido na direção da interoperabilidade semântica. Após a hierarquização das classes, começa-se a viabilizar a construção da semântica, na medida em que é possível utilizar estruturas XML e RDF (*Resource Description Framework*) para representar estas hierarquias em linguagem processável por máquinas. O uso de vocabulários padrão, tais como SKOS (*Simple Knowledge Organization System*) e SCOVO (*The Statistical Core Vocabulary*) além do VCGE, é condição importante para viabilizar o reuso deste trabalho em outros esforços que estão sendo planejados no âmbito do Governo Brasileiro.

Verificamos que o VCGE se constitui em um instrumento dinâmico de organização do conhecimento, podendo ser utilizado para diversos fins, de acordo com os interesses de cada usuário e sofrendo alterações de acordo com as necessidades. Entretanto, notamos indícios, alguns apontados aqui, que ele não conseguirá contemplar as relações semânticas presentes nos dados referentes a um domínio tão complexo como o do Estado Brasileiro.

Por fim, como o VCGE é o vocabulário indicado para representar as temáticas governamentais, acredita-se que os seus responsáveis devam se utilizar de metodologias e teorias propostas por pesquisadores, Instituições Governamentais e demais instrumentos de representação do conhecimento de instituições internacionais para aprimorá-lo, com vista a torná-lo mais próximo da realidade representada e conseqüentemente mais efetivo para o propósito a que se destina.

REFERÊNCIAS

ANUÁRIO ESTATÍSTICO DA PREVIDÊNCIA SOCIAL: AEPS 2011. Brasília, v. 19, 2012.

ANUÁRIO ESTATÍSTICO DE ACIDENTES DO TRABALHO: AEAT 2010. Brasília, v. 1, 2011.

APEL, K.-O. **A transformação da Filosofia**. São Paulo: Loyola, 2000.

BARROS, A. J. S.; LEHFELD, N. A. S. **Fundamentos de metodologia**: um guia para a iniciação científica. 2. ed. São Paulo: Makron Books, 2000.

BERNOIT, G. Toward a critical theoretic perspective in information systems. **Library Quarterly**, v. 72, n. 4, p. 441-471, 2002.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. **Plano de ação da Infraestrutura Nacional de Dados Abertos**. Brasília: Secretaria de Logística e Tecnologia da Informação, 2012.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. **Interoperabilidade semântica**. 2013. Disponível em: <<https://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padrees-de-interoperabilidade/interoperabilidade-semantica/index>>. Acesso em: 17 jul. 2013.

BRETAS, N. et al. A construção da e-PING: situação atual e desafios. In: BRASIL. Ministério do planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. **Panorama da interoperabilidade no Brasil**. Brasília: Ministério do planejamento, Orçamento e Gestão, 2010. Disponível em: <www.eping.e.gov.br>. Acesso em: 21 jul. 2013.

CAMPOS, M. L. A. A problemática da compatibilização terminológica e a integração de ontologias: o papel das definições conceituais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 6., 2005, Florianópolis. **A política científica e os desafios da sociedade do conhecimento**. Florianópolis: ANCIB, 2005. Disponível em: <<http://www.uff.br/ppgci/editais/marialuizaleituras.pdf>>. Acesso em: 17 jul. 2013.

CHAUÍ, M. **Convite à filosofia**. 5a. edição. São Paulo: Editora Ática, 1995.

COSTA, C. F. **Filosofia Analítica**. Rio de Janeiro: Tempo Brasileiro, 1992.

DADOS.GOV.BR. 2013. Disponível em: <<http://dados.gov.br/>>. Acesso em: 12 julho 2013.

ERIKSSON, O. A generic communication model based on Habermas' and Searle's versions of speech act theory. **The Language/Action Perspective**, 1999. Disponível em: <<http://mydocs.strands.de/MyDocs/05663/05663.pdf>>. Acesso em: 17 junho 2013.

FROHMANN, B. A documentação rediviva: prolegômenos a uma (outra) filosofia da informação. **Morpheus**: Revista Eletrônica em Ciências Humanas, Rio de Janeiro, v. 9, n. 14, p. 227-249, 2012. Disponível em: <http://www.unirio.br/morpheusonline/numero14-2012/artigos/frohmann_pt.pdf>. Acesso em: 14 out. 2012.

FONSECA, F.; EGENHOFER, M.; BORGES, K. A. V. Ontologias e interoperabilidade semântica entre SIGs. In: WORKSHOP BRASILEIRO DE GEOINFORMÁTICA, 2., 2000, São Paulo. **Anais...** São Paulo, 2000. Disponível em: <<http://s3.amazonaws.com/academia.edu/documents/31034403/GeoInfo2000.pdf?AWSAccessKeyId=AKIAIR6FSIMDFXPEERSA&Expires=1375925834&Signature=0gXHp88Iw3ldeW14u3e1w9u1h14%3D&response-content-disposition=inline>>. Acesso em: 17 jul. 2013.

GRUPO DE TRABALHO DA INFRAESTRUTURA NACIONAL DE DADOS ABERTOS. **GT INDA**. 2013. Disponível em: <<http://wiki.gtinda.ibge.gov.br/MainPage.ashx>>. Acesso em: 21 jul. 2013.

HJØRLAND, B. Fundamentals of Knowledge Organization. In: FRÍAS, J. A.; TRAVIESO, C. (Ed.). **Trends in Knowledge Organization Research**. Salamanca: Ediciones Universidad de Salamanca, 2003. p.83-116.

HUSSERL, E. **Ideias para uma fenomenologia pura e para uma filosofia fenomenológica**. São Paulo: Ideias & Letras, 2006.

JAPIASSÚ, H.; MARCONDES, D. **Dicionário básico de Filosofia**. 4. ed. Rio de Janeiro: Jorge Zahar Ed., 2006.

LIMA, J. A. O. Interoperabilidade semântica no LexML. In: BRASIL. Ministério do planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. **Panorama da interoperabilidade no Brasil**. Brasília: Ministério do planejamento, Orçamento e Gestão, 2010. Disponível em: <www.eping.e.gov.br>. Acesso em: 21 jul. 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. Disponível em: <<http://dx.doi.org/10.3145/epi.2012.may.03>>. Acesso em: 25 mar. 2013.

MIRANDA, C. M. C. A disseminação de dados governamentais abertos como serviço público – os dados abertos governamentais e a experiência brasileira. In: CONGRESSO INTERNACIONAL SOFTWARE LIVRE E COMÉRCIO ELETRÔNICO, 4., 2011, Brasília. **Dados abertos para a democracia na era digital**. Brasília: Fundação Alexandre de Gusmão, 2011. p. 63-73.

MUCHERONI, M.; SILVA, J. F. M.. A interoperabilidade dos sistemas de informação sob o enfoque da análise sintática e semântica de dados na *Web*. **PontodeAcesso**, Salvador, v. 5, n. 1, p. 03-18, abr. 2011. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/viewArticle/3622>>. Acesso em: 21 jul. 2013.

PATEL, M. et al. Semantic Interoperability in Digital Library Systems. In: **WP5 Forum Workshop: Semantic Interoperability in Digital Library Systems**, DELOS Network of Excellence in Digital Libraries, 2004-09-16, Bath. Bath, UK: University Bath, 2004. Disponível em: <http://opus.bath.ac.uk/35098/4/WP5_Workshop_040916.pdf>. Acesso em: 21 jul. 2013.

RIBEIRO, C. J. S. **Diretrizes para o projeto de portais de informação: uma proposta interdisciplinar baseada na Análise de Domínio e Arquitetura da Informação**. 2008. 298f. Tese (Doutorado em Ciência da Informação). Universidade Federal Fluminense/IBICT, Rio de Janeiro.

RIBEIRO, C. J. S. ; ALMEIDA, R. F. . Dados Abertos Governamentais (Open Government Data): Instrumento para Exercício de Cidadania pela Sociedade. In: **Elmira Simeão, Jorge Henrique Cabral Fernandes, Isa Maria Freire. (Org.). XII Enancib - Políticas de Informação para a Sociedade - Anais**. Brasília: Thesaurus, 2011, p. 2568-2580.

SAYÃO, L. F.; MARCONDES, C. H. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **TransInformação**, Campinas, v. 20, n. 2, p.133-148, maio/ago. 2008.

SOUZA, R. R.; ALVARENGA, L. A *Web Semântica e suas contribuições para a ciência da informação*. **Ci. Inf.**, Brasília, v. 33, n. 1, abr. 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n1/v33n1a16.pdf>>. Acesso em: 21 jul. 2013.

SOWA, J. F. **Knowledge Representation**. Ed. Brooks/Cole. CA. 2000.

W3C. Escritório 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: 28 jun. 2011.

WACHE, H. et al. Ontology-based integration of information - a survey of existing approaches. **In: PROCEEDINGS of IJCAI-01 Workshop: Ontologies and Information Sharing**. Seattle (EUA): [s.n.], 2001. p. 108-117.

WERSIG, G. Information Science: the study of postmodern knowledge usage. **Information Processing & Management**, v. 29, n. 2, p. 229-239, 1993.