

PSO: Ontologia de Domínio para Publicação de Dados sobre Servidores Públicos

Victor Gonçalves¹, Bernadette Farias Lóscio¹

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Cidade Universitária – CEP 50740-560 – Recife, PE – Brasil

{vg3,bfl}@cin.ufpe.br

Abstract. *This paper presents the implementation of a domain ontology intended to represent data about public servants. Aiming to represent an important domain in a standardized way to society, it makes use of consolidated ontologies and proved to be functional in the first analysis.*

Resumo. *Este artigo apresenta a implementação de uma ontologia de domínio destinada a representar os dados sobre servidores públicos. Visando descrever de forma padronizada um domínio importante à sociedade, a mesma faz uso de ontologias consolidadas e se mostrou funcional em primeira análise.*

1. Introdução

Ao disponibilizarem os dados que gerenciam, os governos possibilitam uma maior participação social, transparência e geração de valor social e comercial, motivando a criação de novos produtos e serviços por diferentes setores da sociedade (ATTARD et al., 2016). Isto porque tais dados abrangem diversas áreas e são de grande interesse público. Assim, um dos grandes benefícios obtidos ao se utilizar Dados Abertos está na possibilidade de combiná-los e gerar novos dados e conhecimento antes desconhecidos (OKFN, 2015).

No Brasil, existem duas grandes iniciativas em resposta à necessidade de tornar a informação pública, que são o Portal de Dados Abertos e o Portal da Transparência. No entanto, em levantamento recente, mesmo o país estando à frente dos demais países da América Latina na abertura de dados, ainda possui gargalos quanto: (i) à usabilidade dos dados, ou a dificuldade de se trabalhar com eles, (ii) à disponibilização de download completo e (iii) à localização dos mesmos (OKBR, 2017). Assim, evidencia-se que um importante desafio enfrentado atualmente está relacionado à ausência de padrões e informações adicionais (semântica) associados aos dados. Tais fatores são essenciais à compreensão e manipulação dos mesmos, facilitando a integração dos diversos conjuntos de dados fornecidos pelas entidades governamentais e o reuso desses dados pelos cidadãos.

Entre as estratégias que visam tratar esse problema está a recomendação proposta pelo *World Wide Web Consortium*¹⁸ (W3C), que prevê entre outras práticas o uso de ontologias e tecnologias semânticas para integrar e disponibilizar dados. Tal atividade aumenta o consenso entre os envolvidos, reduzindo redundâncias e ambiguidades, mas também amplia a interoperabilidade e incentiva o reuso dos dados (LÓSCIO et al., 2017).

¹⁸ <https://www.w3.org>

2. PSO

Carossi e Teixeira Filho (2016), ao analisarem os pedidos de acesso à informação encaminhados às Instituições de Ensino Superior por meio do Sistema Eletrônico do Serviço de Informação ao Cidadão (e-SIC), verificaram que o principal tema demandado pelo cidadão está relacionado ao quadro de servidores das instituições. Porém, considerando que muitas dessas informações já se encontram disponíveis em portais públicos do Governo Brasileiro, é possível inferir que tais dados são apresentados de modo ineficiente, levando ao cidadão a recorrer à Lei de Acesso à Informação.

Dessa forma, a *Public Servant Ontology* - PSO visa representar de maneira padronizada os dados relacionados aos Servidores Públicos Federais, atualmente disponibilizados no Portal da Transparência, de modo a facilitar a interoperabilidade e o reuso dos mesmos.

2.1. Desenvolvimento

Para a implementação da PSO, optou-se pela utilização da metodologia NeOn, uma vez que a mesma é baseada em metodologias consolidadas como Methontology, On-To-Knowledge e Diligent mas também em boas práticas e *feedback* de experiências dos integrantes do consórcio¹⁹ idealizador do projeto Neon.

A metodologia Neon prevê nove cenários distintos a fim de compreender os principais contextos de desenvolvimento de ontologias, onde cada cenário contém um conjunto de processos e atividades bem definidos, identificando os atores envolvidos e os artefatos gerados em cada etapa (SUÁREZ-FIGUEROA et al., 2012).

Com base nos principais temas associados à categoria Quadro de Servidores mapeados por Carossi e Teixeira Filho (2016), a saber, (i) quantitativo, (ii) lotação, (iii) carga horária, (iv) função gratificada, (v) remuneração e (vi) afastamento, foi possível extrair Questões de Competências (QC) da base de dados disponibilizada no portal do eSIC²⁰ a fim de compor o Documento de Especificação de Requisitos da Ontologia (ORSD) previsto na Neon. Tais QCs contribuem também na fase de avaliação da ontologia, onde verifica-se a adequação da mesma ao contexto modelado.

Tendo em vista a reutilização de recursos ontológicos especificados na Neon, foram realizadas buscas no repositório LOV²¹, a fim de encontrar vocabulários/ontologias relacionadas ao contexto da PSO. A busca considerou os termos extraídos do ORSD. Assim, foi possível reutilizar as seguintes ontologias:

- *The Organization Ontology*²² – ontologia utilizada para descrever estruturas organizacionais;
- *Friend of a Friend (FOAF)* – ontologia destinada à descrição de pessoas e seus relacionamentos com objetos e outras pessoas;
- *Time Ontology* – ontologia para representação de intervalos e instantes;
- *Simple Knowledge Organization System (SKOS)* – ontologia destinada à

19 <http://www.neon-project.org/nw/Partners>

20 http://www.consultaesic.cgu.gov.br/busca/_layouts/15/DownloadPedidos/DownloadDados.aspx

21 <http://lov.okfn.org/dataset/lov/>

22 <https://www.w3.org/TR/vocab-org/>

esquemas conceituais, tais como listas.

Foi possível implementar uma primeira versão da PSO²³ em *Web Ontology Language* (OWL), padrão definido pelo W3C, conforme apresentado na Figura 1. Em termos gerais, foram adicionados à *The Organization Ontology* as Classes *Remuneration*, *WorkLeave* para representarem respectivamente a remuneração e afastamento dos servidores, e também as propriedades de objeto *leavesDuring*, *memberDuring* e *holdsDuring* para capturarem a dimensão temporal associada às respectivas classes, mas também as propriedades *legalRegime*, *weeklyHours* para representarem o regime de trabalho e carga horária do servidor no vínculo com a instituição.

3. Resultados Preliminares

Com o intuito de avaliar inicialmente a PSO, dados sobre os servidores foram extraídos do Portal da Transparência via protótipo ETL desenvolvido com o auxílio da ferramenta Open Studio for Data Integration²⁴ e mapeados em uma base de dados relacional. Em um segundo momento, o mapeamento dos dados nos conceitos previstos na PSO foi realizado utilizando-se a linguagem R2RML. O extrator bem como o mapeamento utilizados estão disponíveis em repositório público²⁵.

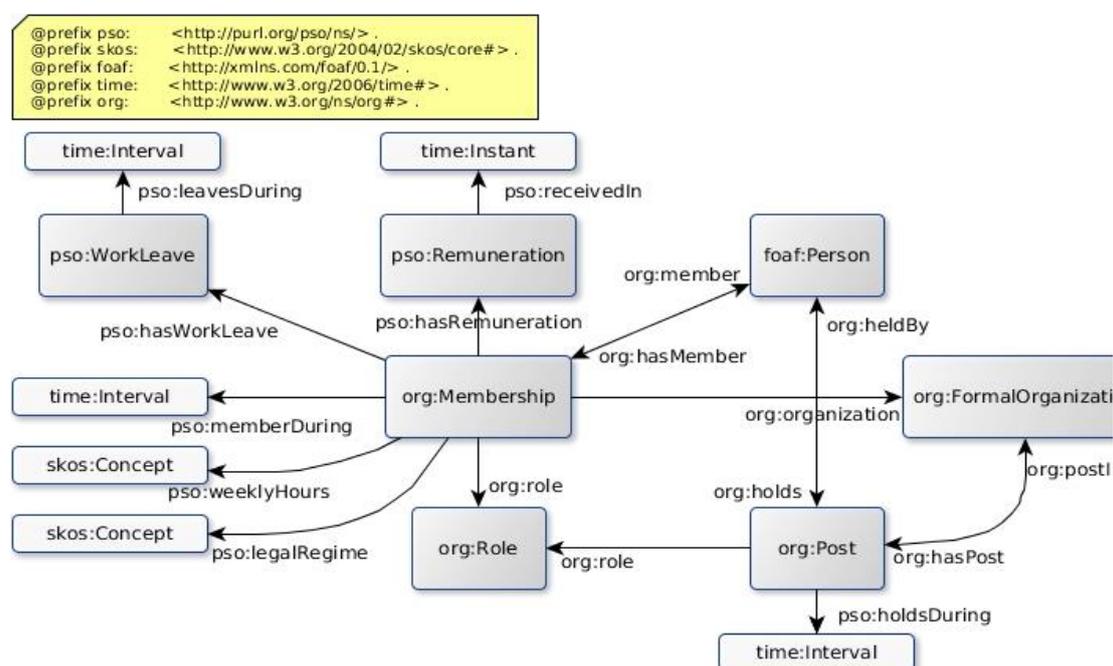


Figura 1: PSO v0.1

Após o mapeamento, realizou-se a triplificação dos dados em *Resource Description Framework* (RDF), disponibilizando-os em um *SPARQL EndPoint* por meio do servidor Apache Jena Fuseki²⁶. Baseando-se nas QCs extraídas do eSIC, foi realizada uma consulta SPARQL relacionada ao quantitativo servidores em um determinado cargo.

23 <http://purl.org/ns/pso/>

24 <https://www.talend.com/products/talend-open-studio/>

25 <https://github.com/hotcyv/mestrado>

26 <http://jena.apache.org/documentation/fuseki2/index.html>

O resultado apresentado pela PSO comparado à respectiva resposta da instituição ao solicitante foi igual, demonstrando inicialmente a capacidade da PSO em responder de forma automatizada e padronizada informações sobre servidores públicos.

Ainda, foi possível verificar também a possibilidade de integração da PSO com a ontologia do Orçamento Público Federal (ARAÚJO et al., 2015), a qual destina-se a representar o domínio do orçamento brasileiro, porém sem considerar a estrutura organizacional das instituições na apresentação dos dados orçamentários. Seria possível obter um maior detalhamento do orçamento das instituições, evidenciando-se, assim, a interoperabilidade de dados por meio da utilização e integração de ontologias.

Assim, futuramente a PSO será submetida à uma avaliação contendo uma gama maior de QCs envolvendo os demais conceitos modelados bem como à uma avaliação considerando a integração da mesma com a ontologia do orçamento.

Referências

- ATTARD, Judie et al. Value Creation on Open Government Data. In: 2016 49th Hawaii International Conference on System Sciences (HICSS). IEEE, 2016. p. 2605-2614.
- CAROSI, D. F. , TEIXEIRA FILHO, J. G. A. Uma Análise dos Pedidos de Acesso à Informação Encaminhados a uma Instituição de Ensino Superior. Revista Gestão.Org, v. 14, Edição Especial, 2016. <http://www.revista.ufpe.br/gestaoorg>.
- LÓSCIO, Bernadette F., BURLE, C., CALEGARI, N.. Data on the Web Best Practices. W3C Recommendation, 2017. Disponível em: <https://www.w3.org/TR/dwbp/>.
- OKBr, Open Knowledge Brasil e Fundação Getúlio Vargas. Índice de Dados Abertos para o Brasil. Disponível em: <http://dapp.fgv.br/wp-content/uploads/2017/04/IndiceDadosAbertosBrasil2017.pdf>. 2017
- OKFN, Open Knowledge Foundation. The Open Data Handbook. Disponível em: <http://opendatahandbook.org/guide/en/>. 2015.
- SUÁREZ-FIGUEROA, Mari Carmen; GOMEZ-PEREZ, Asuncion; FERNANDEZ-LOPEZ, Mariano. The NeOn methodology for ontology engineering. In: Ontology engineering in a networked world. Springer Berlin Heidelberg, 2012.
- ARAÚJO, L. S. de Oliveira, Santos, M. T. e SILVA, D. A. 2015. The Brazilian federal budget ontology: a semantic web case of public open data In Proceedings of the 7th International Conference on Management of computational and collective intelligence in Digital EcoSystems.