

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DAINF - DEPARTAMENTO ACADÊMICO DE INFORMÁTICA
CURSO DE BACHARELADO EM SISTEMAS DE
INFORMAÇÃO

ESTER BONDEZAN COLAÇO DE LIMA

**ESTUDO DE VIABILIDADE DE REÚSO DA ONTOLOGIA
PLP (*Patient Life Path*) PARA
INTEGRAÇÃO SEMÂNTICA DE DADOS ABERTOS
DA ÁREA DA SAÚDE**

TRABALHO DE CONCLUSÃO DE CURSO

CURITIBA

2019

ESTER BONDEZAN COLAÇO DE LIMA

**ESTUDO DE VIABILIDADE DE REÚSO DA ONTOLOGIA PLP
(*Patient Life Path*) PARA
INTEGRAÇÃO SEMÂNTICA DE DADOS ABERTOS
DA ÁREA DA SAÚDE**

Proposta de Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Sistemas de Informação da Universidade Tecnológica Federal do Paraná, como requisito parcial para a obtenção do título de Bacharel.

Orientadora: Prof.^a Rita Cristina Galarraga Berardi
DAINF - Departamento Acadêmico de
Informática - UTFPR

CURITIBA
2019



Ministério da Educação
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
Câmpus Curitiba
Diretoria de Graduação e Educação Profissional
Departamento Acadêmico de Informática
Coordenação do Curso de Bacharelado em Sistemas de Informação



TERMO DE APROVAÇÃO

“ESTUDO DE VIABILIDADE DE REÚSO DA ONTOLOGIA PLP (Patient Life Path) PARA INTEGRAÇÃO SEMÂNTICA DE DADOS ABERTOS DA ÁREA DA SAÚDE”

por

Ester Bondezan Colaço de Lima

Este Trabalho de Conclusão de Curso foi apresentado como requisito parcial à obtenção do grau de Bacharel em Sistemas de Informação na Universidade Tecnológica Federal do Paraná - UTFPR - Câmpus Curitiba. A aluna foi arguida pelos membros da Banca de Avaliação abaixo assinados. Após deliberação a Banca de Avaliação considerou o trabalho _____.

| | |
|--|---|
| <p>_____</p> <p>Profa. Rita Cristina Galarraga Berardi (Presidente - UTFPR/Curitiba)</p> | <p>_____</p> <p>Prof. Alexandre Graeml (Avaliador(a) 1 – UTFPR/Curitiba)</p> |
| <p>_____</p> <p>Josiane Michalak Hauagge Dallagnol (Avaliador(a) 2 – UTFPR/Curitiba)</p> | <p>_____</p> <p>Profa. Leyza Baldo Dorini (Professora Responsável pelo TCC – UTFPR/Curitiba)</p> |
| <p>_____</p> <p>Prof. Marcelo Mikosz Gonçalves (Coordenador do curso de Bacharelado em Sistemas de Informação – UTFPR/Curitiba)</p> | |

“A Folha de Aprovação assinada encontra-se na Coordenação do Curso.”

RESUMO

Este trabalho visa identificar as dificuldades que podem ser encontradas no reúso da ontologia PLP criada no trabalho anterior pelos autores Auceli e Garcia (2019, no prelo). Baseado na utilização de bases de dados abertos de Curitiba, os autores realizaram um trabalho de fornecer uma integração semântica para realizar o mapeamento da trajetória de pacientes que recebem atendimento em unidades públicas de saúde, desde tratamentos iniciais nas Unidades de Pronto Atendimento, passando por internamento em hospitais, mas que foram a óbito. Para o reúso desta ontologia, neste trabalho são utilizadas novas bases, contendo dados do mesmo domínio, mas de diferente localidade, neste caso, a cidade de São Paulo – SP. Mesmo com uma ontologia no mesmo domínio (de saúde) alguns desafios podem ser encontrados ao reusá-la ainda que seguindo guias de Melhores Práticas para Reúso. O objetivo deste trabalho é identificar e explorar esses desafios destacando as vantagens de reúso e como as Melhores Práticas podem auxiliar neste processo, analisando possíveis alterações e melhorias no processo de construção de uma ontologia reutilizável.

Palavras-chave: Web-Semântica. Ontologia. Dados abertos governamentais. *Linked Data*. Integração semântica.

ABSTRACT

This paper aims to identify the difficulties that may be found in the reuse of PLP ontology created in the previous work by authors Auceli and Garcia (2019, forthcoming). Based on the use of open databases in Curitiba, the authors performed a work to provide a semantic integration to map the trajectory of care patients in public health units, from initial treatments in the Emergency Care Units to hospitalization, but coming to die. For the reuse of this ontology, this paper uses new databases containing data from the same domain, but from a different location, in these case, the city of São Paulo – SP. Even with an ontology in the same domain (health) some challenges can be met by reusing it even by following reuse Best Practice guides. The aim of this paper is to identify and explore these challenges by highlighting the reuse advantages and how Best Practices can assist in this process by analyzing possible changes and improvements in the process of building a reusable ontology.

Palavras-chave: Web-Semantics. Ontology. Government Open Data. Linked Data. Semantic Integration.

LISTA DE TABELAS

| | |
|---|----|
| Figura 1 – Classe Diagnosis e suas propriedades | 21 |
| Figura 2 – Subclasse Hospital e suas propriedades | 22 |
| Figura 2 – Classe Person e suas propriedades | 22 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 – Comparação entre bases de dados deste trabalho e do trabalho de Auceli e Garcia (2019, no prelo) | 19 |
| Tabela 2 – Ontologia PLP - Classes e Subclasses | 20 |
| Tabela 3 – Ontologia PLP - Propriedades do Objeto | 21 |
| Tabela 4 – Ontologia PLP - Propriedades de Dados | 22 |
| Tabela 5 – Tabela de filtragem de dados por base utilizada | 27 |
| Tabela 6 – Tabela de conversão de campos Raça/Cor | 28 |
| Tabela 7 – Tabela de mapeamento de triplas: Classe Person | 30 |
| Tabela 8 – Tabela de mapeamento de triplas: Classe Diagnosis | 30 |
| Tabela 9 – Tabela de mapeamento de triplas: Classe Hospital | 30 |
| Tabela 10 – Tabela de benefícios decorrente da adoção das Melhores Práticas | 33 |
| Tabela 11 – Tabela de definições de Melhores Práticas adotadas neste estudo | 33 |

SUMÁRIO

| | |
|---|-----------|
| 1. INTRODUÇÃO | 8 |
| 2. REFERENCIAL TEÓRICO | 10 |
| 2.1. <i>LINKED DATA</i> E WEB SEMÂNTICA | 11 |
| 2.1.1. RDF e RDFS | 13 |
| 2.2. ONTOLOGIA | 14 |
| 2.3. TRABALHOS RELACIONADOS | 16 |
| 3. METODOLOGIA | 18 |
| 3.1. SELEÇÃO DAS BASES DE DADOS PARA INTEGRAÇÃO | 18 |
| 3.2. ANÁLISE E ATUALIZAÇÃO DA ONTOLOGIA A SER REAPLICADA: PLP | 19 |
| 3.3. PREPARAÇÃO DOS DADOS ABERTOS | 25 |
| 4. RESULTADOS | 31 |
| 4.1. RESULTADO E DISCUSSÃO SOBRE O REÚSO | 31 |
| 5. CONCLUSÃO | 36 |
| REFERÊNCIAS | 39 |
| ANEXOS | 41 |
| ANEXO A - LAYOUT DO ARQUIVO DE PRODUÇÃO AMBULATORIAL: PAUFAAMM.DBF | 42 |

1. INTRODUÇÃO

A disponibilização de dados abertos é algo bem difundido em diversas áreas, mas para que os dados possam ser utilizados de forma integrada, é necessária uma transformação que possibilite que os dados se transformem em informação. A área de *Linked Data* (Dados Conectados) visa promover padrões de interoperabilidade de dados abertos (Rocha, 2012) que possibilite esse uso integrado. A estrutura de dados do *Linked Data* permite a inserção desses dados de forma estruturada na *web*, de forma que podem, não apenas ser consultados, mas conectado com outros tipos de dados com características semelhantes, tornando o universo onde estão inseridos, um multiverso de informações acessíveis.

Quanto melhor definida uma ontologia, a qual identifica e conecta os dados, mais bases ela poderá alcançar e mais dados poderão ser interpretados e modelados. A reutilização de uma ontologia é a iniciativa de tornar reutilizável uma semântica criada e até mesmo adaptá-la, se necessário. A reutilização de uma ontologia é importante para que se possa pôr em prática o principal objetivo do *Linked Data*: dados abertos são disponibilizados para que possam ser acessados e reanalisados diversas vezes e por diversos ângulos. Esse conceito preza em manter maior visibilidade e possibilitar a circulação de dados, que até então se encontrava restrita em fontes fora do alcance público, em silos de dados. Com a ampla visibilidade de dados, torna-se possível aplicar uma ontologia para um determinado escopo com bases de dados de diversos locais que tenham os mesmos tipos de dados que possam ser referenciados.

Conforme citado por Rocha (2012), a Web Semântica é a solução para a integração de incontáveis quantidades de metadados de diversas fontes, porém, se para cada novo estudo for implementada uma nova ontologia para interpretação de dados do mesmo escopo, a Web Semântica não estará alcançando seu objetivo de utilizar dados abertos para que não haja múltiplos registros semelhantes em diversas bases espalhadas.

Este trabalho utiliza dados na área de saúde, tendo como motivação compreender os desafios relacionados ao reúso de uma ontologia já criada e utilizada anteriormente. Os autores Auceli e Garcia (2018, no prelo), realizaram o trabalho de “INTEGRAÇÃO SEMÂNTICA DE BASES DE DADOS ABERTOS DE SAÚDE DE CURITIBA: Um estudo de caso com atendimentos com desfecho em óbito”. Os

autores realizaram uma integração com dados de bases de saúde de Curitiba para que se possa obter informações mais ricas com base no relacionamento de dados. Os dados utilizados pelos autores foram obtidos de três fontes diferentes, conectando as características comuns para obter início e fim de informação sobre a trajetória de pacientes. O caminho mapeado foi desde a realização de atendimentos a pacientes em unidades de pronto atendimento (UPA), na cidade de Curitiba, que foram internados para tratamento e que, por fim, vieram a óbito.

Para o reúso serão considerados dados da cidade de São Paulo/SP e para isso será necessário substituir uma das bases já que é específica do contexto municipal. Por meio da reutilização da ontologia, busca-se explorar a utilidade da semântica fora dos parâmetros para o qual ela foi criada, verificando, assim, se a ontologia criada pode ser utilizada em outras fontes de dados do mesmo domínio. Caso não, procura-se identificar quais as dificuldades encontradas quando se deseja reutilizar uma ontologia. Além disto neste processo serão consideradas as melhores práticas para reúso, apontadas pela Convenção W3C.

Por fim, serão analisadas e discutidas vantagens e os desafios encontrados no reúso da ontologia PLP, o que pode ser aprimorado e a validade de reaplicar a ontologia em bases com informações diferentes, porém de mesmo escopo das bases utilizadas pelos autores na ontologia PLP.

2. REFERENCIAL TEÓRICO

Para a difusão de dados na *web*, a utilização de bancos de dados relacionais não é eficiente, pois estes não possuem a capacidade de integração de dados (Zaidan e Bax, 2013). Os dados abertos conectados (*Linked Open Data*) são uma forma de publicar e relacionar conjuntos de dados (*datasets*) na *web*.

O *Linking Open Data* é um projeto aberto e colaborativo de responsabilidade do Consórcio W3C¹ – *World Wide Web Consortium*, uma comunidade internacional com membros que trabalham juntos para desenvolver padrões da *web*, liderada pelo inventor da *web*, Tim Berners-Lee, sua missão é levar a *web* a todo o seu potencial na publicação de dados.

Conforme Araújo e Souza (2011), a W3C busca identificar *datasets* existentes com licenças abertas e transformá-los em modelo de padrões de dados com representação semântica e com os princípios de *Linked Data (Dados Conectados)*. A W3C também possui guias de publicação de *Linked Open Data*, chamadas “Melhores Práticas”, que em uma série de documentos procura trazer práticas facilitadoras no desenvolvimento e publicação de dados abertos.

Nos últimos anos, os governos de todo o mundo têm realizado publicação de dados governamentais na *web* com o objetivo de atingir maior transparência em suas ações, e para facilitar o apoio da sociedade em iniciativas governamentais (*World Wide Web Consortium*, 2014). Porém dados abertos não são necessariamente conectados, assim como há dados que são conectados, mas não são abertos. Segundo a *World Wide Web Consortium* (2014), o *Linked Data* fornece um mecanismo simples para combinar dados de várias fontes na *web*, com iniciativas abertas de transparência do governo por meio do uso de padrões internacionais da *web* para publicação, disseminação e reutilização de dados estruturados.

Inúmeras pesquisas estão sendo realizadas na exploração da Web Semântica, considerando boas práticas de reutilizar termos conhecidos de vocabulários, como no trabalho de Araújo e Souza (2011), que para a criação de sua base de dados, precisaram coletar dados de diversas fontes, integrando e agrupando os tipos de

¹ <https://www.w3.org/Consortium/>

dados para, após detectar as descrições de cada entidade assimilando com as bases comparativas, pudessem transformar as informações em *Linked Data*.

A acessibilidade dos dados abertos permite a combinação de fontes de dados diferentes, produzindo novos significados para a mesma informação e fornecendo desse modo, informações realmente úteis aos cidadãos (Araújo e Souza, 2011).

Assim como no trabalho apresentado por Zaidan e Bax (2013), a integração de dados a partir de múltiplas fontes de dados públicas foi realizada para a identificação da viabilidade e identificação de quais problemas surgem e quais adaptações são necessárias ao reaplicar uma ontologia que pode ser utilizada para traduzir, compreender e unir dados de diversas fontes.

2.1. LINKED DATA E WEB SEMÂNTICA

O termo “*Linked Data*”, ou “Dados Conectados”, é uma definição da publicação de dados de forma que são legíveis (interpretáveis) por máquinas, tendo significados explicitamente definidos e que estão obrigatoriamente ligados a outros conjuntos de dados, que por sua vez possam estar ligados a conjuntos externos de dados, segundo Araújo e Souza (2011). Como uma forma de agregação de valor a dados de sistema, se reflete em não haver dados em ambientes isolados. *Linked Data* refere-se a uma *web* que torna os dados mais úteis ao poder conectá-los em relacionamentos que diminuem as barreiras de conhecimento e utilização da informação (Zaidan e Bax, 2013).

Os dados abertos têm o foco em tornar o dado reutilizável, podendo ser redistribuído por qualquer pessoa, segundo descrevem Siqueira *et al.* (2016), gerando assim informações para melhores tomadas de decisões, promovendo competitividade e inovação, permitindo que pessoas e máquinas possam trabalhar de modo mais eficiente juntos. Rocha (2012) também ressalta que os dados abertos proporcionam publicações de baixo custo, ampliando o acesso a conteúdos que eram de circulação limitada.

Segundo Siqueira *et al.* (2016), *Linked Data* possibilita o trabalho mais eficiente entre humanos e máquinas, auxilia a revelar o que antes estava escondido na “*Web de documentos*” e agora se torna acessível graças a esse padrão de estruturação e compartilhamento de dados. De acordo com Bizer *et al.* (2011), quanto mais um dado

estiver ligado a outro, maior valor e utilidade ele terá, sendo que *Linked Data* são dados publicados de forma estruturada, sob um conjunto de melhores práticas para publicação e conexão dos dados na *web*. Essa prática tem alcançado diversos domínios conectando bases de dados globais por meio da “*Web de Dados*”, que contém bilhões de dados de entrada, aumentando cada vez mais o número de fontes de dados.

Marcondes (2012) refere-se a dados conectados como uma navegação orientada por *links* que permite uma transição entre diferentes fontes de informação, que possibilitam interligar acervos em arquivos, bibliotecas e museus digitais por meio de tecnologias da Web Semântica. Marcondes (2012) traz ainda a ideia de que os *links* convencionais da *web* são apenas meios que programas de navegadores utilizam, transitando o acesso de um recurso a outro, sem explicação do real significado entre estes recursos. Porém os *links* semânticos podem ser processados de forma mais rica, tendo um significado semântico de ligação entre ambos os recursos no âmbito informacional, semântico, cultural, cognitivo, econômico ou científico.

O conceito de “*Linked Data*” é visto como a *web* de dados, que são armazenados em bases de dados, interligados através da rede, referenciando uma à outra, segundo Rocha (2012). Segundo Zaidan e Bax (2013), os criadores de *Linked Data* estabelecem princípios para esclarecer a estrutura desses dados, que devem ser: abertos - podendo ser acessados por qualquer aplicação; modulares - sem a necessidade de planejamento prévio de acordo de dados; e escaláveis - possibilitando a adição de novos dados relacionados. Ressaltam também que o acesso a esses dados é realizado por meio de um vocabulário padrão que será especificado em cada ontologia, e que existem vários domínios de naturezas diversas com significativos conjuntos de dados, como por exemplo o DBPedia, *dataset* da Wikipédia, que é uma das bases de dados abertos e conectados que pode ser acessada livremente por qualquer aplicação.

Web Semântica é a possibilidade de definir expressividade para os dados de acordo com Zaidan e Bax (2013). É um instrumento que fornece sentido lógico e semântico de dados aos computadores. Web Semântica ainda é descrita pelos autores como uma extensão da *web* atual de textos e documentos. A *web* atual, chamada de Web Sintática por Marcondes (2012), pois trabalha com padrões sintáticos de “formas” que, ao realizar buscas em programas, procura por padrões de

textos idênticos ao fornecido como parâmetro de pesquisa, não permitindo assim a interpretação do que se procura. Marcondes (2012) ainda diz que a proposta da Web Semântica descreve uma *web* diferente da *web* atual, com conteúdos exibidos de forma que possam ser lidos por pessoas e ainda ter uma semântica interpretável por programas, os chamados agentes de software, não mais se utilizando de conteúdos estruturados.

Enquanto a Web Semântica, ou “Web de Dados”, tem como objetivo o processamento dos dados de forma que possam ser interpretados naturalmente por máquinas, *Linked Data* utiliza-se dessas informações resultantes para trazer sentido aos dados. Bizer *et al.* (2011) dizem que a publicação de *Linked Data*, feita por inúmeros grupos e indivíduos, foi o que contribuiu para a população da Web Semântica. A Web Semântica é a plataforma de representação do *Linked Data*, como explica Rocha (2012), composta por triplas que representam os dados na linguagem RDF (*Resource Description Framework*).

Existem quatro princípios citados por Araújo e Souza (2011) que caracterizam o *Linked Data* e que devem ser aplicados para a Web Semântica crescer. O primeiro é a utilização de URI (*Uniform Resource Identifier*) para nomes de recursos, um identificador exclusivo para cada item de interesse na *web*, chamado de “recurso”. O segundo é a utilização de URIs HTTP para que esses identificadores possam ser pesquisados por pessoas, pois por meio do protocolo HTTP pode-se acessar um URI na *web* a fim de obter informações sobre qualquer recurso identificado. O terceiro princípio é o fornecimento de informações úteis no acesso de um URI, utilizando padrões estruturados como a linguagem RDF e o protocolo SPARQL, que é a forma de realizar as consultas nos dados RDF. O quarto princípio é a inclusão de *links* de um URI para outro URI, para que pessoas possam encontrar mais recursos originando-se do recurso inicial, tecendo assim uma teia de inúmeros dados conectados. Araújo e Souza (2011) ainda definem as URIs, RDF e HTTP como as principais tecnologias utilizadas na disponibilização de *Linked Data*.

2.1.1. RDF e RDFS

Rocha (2012) cita que os dados na linguagem RDF, em suas estruturas de triplas, são os metadados que compõem os *datasets* (armazenamento de dados

conectados), termo utilizado na recomendação de melhores práticas de compartilhamento e conexão de dados. Os *links* RDF permitem que as aplicações naveguem entre as fontes de dados e acessem dados adicionais conectados, dizem Araújo e Souza (2011), sendo que para estarem inseridas na Web de Dados, as fontes de dados devem ter *links* RDF definidos para relacionarem-se.

Zaidan e Bax (2013) dizem que o RDF pode ser considerado um modelo de dados com uma semântica simplificada no tratamento de metadados. Araújo e Souza (2011) definem um recurso RDF, na representação como uma série de triplas, estruturada em: sujeito, predicado e objeto. Os autores explicam que:

O sujeito tem o recurso (URI) o qual será escrita a sentença. O predicado ou propriedade, que também é um recurso (URI), representa o relacionamento entre sujeito e um objeto. Já o objeto é o recurso ou um literal que se relaciona com o sujeito. (ZAIDAN E BAX, 2013, p.47)

Segundo Rocha (2012), URIs identificam e interligam os dados, referenciando um ao outro. Enquanto a linguagem RDF tem seu significado formalmente especificado por meio de uma ontologia, que por sua vez é expressa na linguagem OWL (*Web Ontology Language*), que, segundo Schiessl (2007), é uma forma de representar um vocabulário, visão de mundo, indivíduos ou grupos.

Bizer *et al.* (2011) explicam uma estrutura de dados em RDF como o seguinte exemplo: duas pessoas A e B identificadas por um URI único estão relacionadas pelo fato de se conhecerem. Então podemos definir que "A conhece B". Sendo "A" o sujeito da estrutura, "conhecer" o predicado, e "B" o objeto. Outro exemplo demonstrado é afirmação de que "Uma pessoa C é o autor da obra D", sendo assim, a definição RDF seria: "C criou D".

2.2. ONTOLOGIA

A *web* atualmente é composta por páginas de texto, construídos com utilização de palavras que podem ter sentidos distintos em diferentes contextos, pois são textos elaborados por pessoas, afirma Schiessl (2007). Ainda segundo o autor, a recuperação dessas informações necessita ser feita de maneira automática, auxiliada por computadores capazes de tratar as ambiguidades que as linguagens humanas trazem.

Ontologia é um tipo especial de objeto de informação ou artefato computacional, como definem Guarino *et al.* (2009). De forma mais detalhada, os autores ainda esclarecem que uma ontologia trata de uma teoria lógica, projetada para capturar os modelos pretendidos que correspondem a uma certa conceitualização e para excluir os modelos não correspondentes. O resultado será uma especificação aproximada de uma teoria lógica projetada para explicar o significado pretendido de um vocabulário usado por uma linguagem lógica (GUARINO *et al.*, 2009).

A ontologia pode ser definida também como a forma de entender a realidade e processá-la, pois, as ontologias automatizam a interpretação de significado da informação contida em documentos, como sendo parte integrante de um sistema formal, um meio representativo do conhecimento (SCHIESSL, 2007).

Conforme explicam Zaidan e Bax (2013), a ontologia contém a especificação da forma de acesso a *Linked Data* por meio dos vocabulários de um domínio específico. As ontologias especificam o significado das triplas, representadas na linguagem RDF e expressas na linguagem OWL (*Web Ontology Language*). São o RDF e OWL que adicionam semântica aos recursos sem fazer suposições sobre suas estruturas (ROCHA, 2012).

Pan *et al.* (2006) explicam que, conforme definido pelo padrão W3C, uma ontologia OWL consiste em um conjunto de axiomas, podendo ser do tipo classe, propriedade ou indivíduo. O padrão W3C traz definições de que o OWL vai além de linguagens como RDF e XML. Ela tem mais capacidade de expressar o significado e a semântica do dado. Escrita como um conjunto de documentos cada uma cumpre uma finalidade diferente ou atende um público diferente. Pan *et al.* (2006) também citam que uma ontologia pode importar outras ontologias OWL, com ajuda de anotações de importação na sua linguagem, permitindo a importação semântica de classes, propriedades e indivíduos da ontologia de origem. Dessa forma é possível unir duas ou mais ontologias, permitindo a utilização de informações de ambas as ontologias na mesma consulta.

Uma ontologia deve permitir a definição de novos termos baseados em vocabulários existentes, possibilitando que novos termos sejam definidos para usos específicos de forma que não se necessite de uma revisão das definições pré-existentes. Segundo Schiessl (2007), uma ontologia é a maneira de dividir a realidade em partes menores, com objetivo de entendê-la e processá-la da melhor forma. Uma

ontologia é uma forma de automatizar a interpretação de significados, fornecendo definições a um vocabulário para representar um conhecimento. Segundo Falbo *et al.* (2007), ontologia é também definida como um artefato de engenharia que constrói um vocabulário de termos, organizados em conjuntos de axiomas e utilizados para criar novas relações ou restringir interpretações, dentro do seu contexto.

O principal objetivo da captura de conhecimento é permitir o seu reuso e compartilhamento, por isso a reutilização de ontologias é uma abordagem a se considerar (MARTINS, 2009). Segundo Bontas *et al.* (2005), o reuso de ontologia é recomendado por padrão em metodologias atuais e boas práticas, como fator essencial na entrega de melhoria de custos e qualidade de ontologias, o que deve reduzir custos de engenharia além da prevenção de retrabalhos, pois a construção de uma ontologia demanda tempo e trabalho intensivo (DORAN *et al.*, 2007).

2.3. TRABALHOS RELACIONADOS

No trabalho de Auceli e Garcia (2019, no prelo), foi apresentado o problema de integração semântica de dados provenientes de bases de dados abertos, no domínio da área da saúde. Utilizou-se de *frameworks* e outras ferramentas para a limpeza de dados encontrados nas bases para que pudessem ser estruturados de forma a possibilitarem a conexão das bases.

Na integração das bases, Auceli e Garcia (2019, no prelo) criaram tipos de perfis de pacientes identificados por um conjunto de colunas que os categorizam, inserindo um indicador (UID) para realizar a comparação de cada perfil entre as bases. Criaram então uma ontologia intitulada "*Patient Life Path*" (PLP), onde foram mapeados os dados encontrados nos *datasets* em estruturas de classes, propriedades de objeto e propriedades de dados, disponibilizados na ontologia. Após esse mapeamento, os dados foram transformados em triplas RDF, possibilitando assim consultas por meio de queries SPARQL, que é a linguagem para consulta a dados conectados (*Linked Data*), para análise de dados.

A possibilidade da reutilização e compartilhamento da informação é o principal objetivo da captura de conhecimento, segundo Martins (2009) em sua tese de mestrado, a qual tem por objetivo geral propor uma organização e representação do conhecimento por meio de ontologias. A autora ainda cita que para desenvolver

conhecimento para a reutilização de ontologias, é necessário que haja modelos para capturar informações e a ontologia pode ser utilizada para organizar o conhecimento, facilitando o acesso e entendimento da informação e permitindo sua reutilização.

Em sua tese, Martins (2009) ainda traz a problemática da falta de padronização na modelagem e representação de ontologias, que é considerada um dos principais obstáculos para o reuso da ontologia. Ressalta também que embora os recursos de software tenham sido desenvolvidos para o reuso, poupando esforço de adaptação, na prática, na grande maioria dos casos os recursos acabam sendo reconstruídos do zero.

3. METODOLOGIA

Para este trabalho, foi definido como escopo dados da área de saúde que abrangem a cidade de São Paulo, no mesmo domínio da ontologia criada pelos autores Auceli e Garcia (2019, no prelo).

3.1. SELEÇÃO DAS BASES DE DADOS PARA INTEGRAÇÃO

As bases de dados utilizadas por Auceli e Garcia (2019, no prelo) são todas de dados da área de saúde na cidade de Curitiba, visando rastrear, em ocorridos de 2017, características e o destino de um paciente desde seu primeiro atendimento até que venha a óbito. Na busca do rastreamento de perfis de pacientes, foram encontrados e extraídos dados de 3 fontes diferentes.

Conforme pode ser visto na Tabela 1, a primeira base utilizada foi extraída do *site* da prefeitura de Curitiba², contendo dados governamentais que são de domínio público. Trata-se da base do Sistema E-Saude que contém dados referentes ao perfil de atendimentos a pacientes realizados em Unidades de Pronto Atendimento (UPA) da cidade de Curitiba. Outras duas bases foram selecionadas por aqueles autores, ambas extraídas do *site* do DATASUS³ (Departamento de Informática do Sistema Único de Saúde), são elas: SIHSUS, com informações de Sistemas Hospitalares do SUS, e SIM, com dados do Sistema de Mortalidade. Ambas apresentam as características particulares dos pacientes, informações das unidades de atendimento e registro de datas de ocorrências, como a data de óbito na base SIM.

Neste trabalho os dados extraídos são do ano de 2017, o mesmo ano de referência do trabalho de Auceli e Garcia (2019, no prelo). Alterou-se, porém, a localização da ocorrência dos dados, não mais fixado à cidade de Curitiba, mas com o foco em dados da cidade de São Paulo, para que se possa analisar se a utilização de cidades diferentes pode influenciar de alguma forma na reaplicação da ontologia.

A base de dados SIASUS-PA (Sistema de Informação Ambulatorial do SUS - Procedimento Ambulatorial), que pode ser encontrada no *dataset* do site

² <https://www.curitiba.pr.gov.br/dadosabertos>

³ <http://www2.datasus.gov.br/DATASUS>

governamental DATASUS, foi escolhida por conter dados de procedimentos ambulatoriais com registro de informação referente ao paciente atendido e à unidade de saúde que realizou o atendimento.

A base SIHSUS foi mantida neste trabalho por conter dados que podem ser diretamente associados a internamentos de pacientes da base SIASUS. A base SIM foi mantida por se tratar da informação final buscada no trabalho, pacientes que chegam a óbito após atendimento em unidades de saúde.

Tabela 1 – Comparação entre bases de dados deste trabalho e do trabalho de Auceli e Garcia (2019, no prelo)

| Bases usadas em Auceli e Garcia (2019, no prelo) | Bases selecionadas neste trabalho |
|---|---|
| E-saúde - dados de atendimento em UPA referentes a Curitiba/PR | SIASUS - dados de procedimentos ambulatoriais referentes a cidade de São Paulo |
| SIHSUS - dados de internamento hospitalar referentes a Curitiba/PR | A mesma base, porém para dados referentes a cidade de São Paulo |
| SIM - dados de registro de óbito referentes a Curitiba/PR | A mesma base, porém para dados referentes a cidade de São Paulo |

Fonte: Autoria própria (2019).

3.2. ANÁLISE E ATUALIZAÇÃO DA ONTOLOGIA A SER REAPLICADA: PLP

A ontologia PLP criada pelos autores Auceli e Garcia (2019, no prelo), é constituída de 3 classes, 2 subclasses e 3 propriedades de objeto. Neste trabalho, houve a necessidade da criação de uma segunda versão da ontologia PLP, por motivos que serão especificados na Seção 4.3. Na Tabela 2 é possível observar os elementos presentes na versão original de PLP e aqueles que foram mantidos na versão 2.0 da PLP. Além disso, é possível ver a descrição do elemento e a descrição da alteração (quando necessária) na nova coluna “Alterações na PLP v2.0”.

Tabela 2 – Ontologia PLP - Classes e Subclasses.

| Tipo | Nome | Descrição | Coluna de | Alterações na PLP |
|-------------|-------------|------------------|------------------|--------------------------|
|-------------|-------------|------------------|------------------|--------------------------|

| | | | mapeamento PLP | v2.0 |
|-----------|-----------|--|---|---|
| Classe | Person | Especificação de uma pessoa instanciada e mapeada por meio do UID. | Mapeada pelo UID (<i>hash</i> das colunas, “Sexo”, “Data de Nascimento” e “Cidade”) | Mapeada pelo UID2 (<i>hash</i> das colunas “Sexo”, “Idade” e “Cidade”) |
| Classe | Diagnosis | Especificação de um diagnóstico, instanciada pelo código de diagnóstico (CID). | Mapeada pela coluna <i>codigoCid</i> no E-Saude, <i>cid</i> no SIHSUS e <i>cidBasico</i> no SIM | Alteração do mapeamento da base E-saude para a base SIASUS-PA. |
| Classe | Place | Classe genérica para ser utilizada nas subclasses UPA e Hospital. | | |
| Subclasse | Hospital | Subclasse de Place, que especifica um hospital, instanciada pelo CNES do Hospital, mapeada na base SIHSUS. | Mapeamento para as colunas <i>cnes</i> , <i>cnj</i> , <i>municipioestabelecimento</i> | |
| Subclasse | Upa | Subclasse de Place, que especifica uma unidade atendimento | Mapeada para as colunas <i>nomeunidade</i> , <i>codigounidade</i> , <i>nomeunidade</i> . | Não mapeada, dados de atendimentos iniciais serão obtidos na base SIASUS-PA, cujo atendimento é feito em unidades hospitalares. |

Fonte: Autoria própria (2019).

As 3 propriedades de objetos, ou seja, as propriedades que fazem ligação entre classes, foram mantidas na versão 2.0 da ontologia, alterando-se apenas os dados mapeados, como apresentado na Tabela 3.

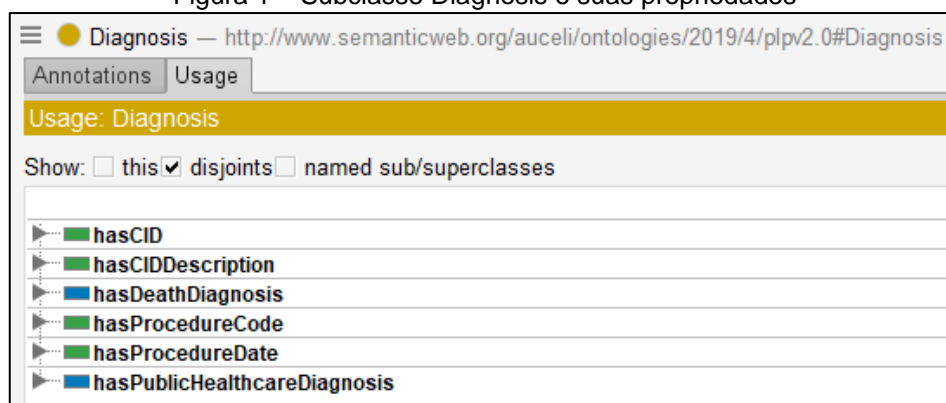
Tabela 3 – Ontologia PLP - Propriedades do Objeto.

| Propriedades do objeto | Descrição | Coluna de mapeamento PLP | Alteração de mapeamento PLP v2.0 |
|-------------------------------|---|---------------------------------|----------------------------------|
| hasDeathDiagnosis | Relacionamento entre a classes Person e Diagnosis | CAUSABAS - base SIM | Não alterado |
| hasHospitalDiagnosis | Relacionamento entre a classes Person e Diagnosis | DIAG_PRINC - base SIHSUS | Não alterado |
| hasPublicHealthcare-Diagnosis | Relacionamento entre a classes Person e Diagnosis | Código do CID - base E-Saude | PA_CIDPRI (CID) - base SIASUS |

Fonte: Autoria própria (2019).

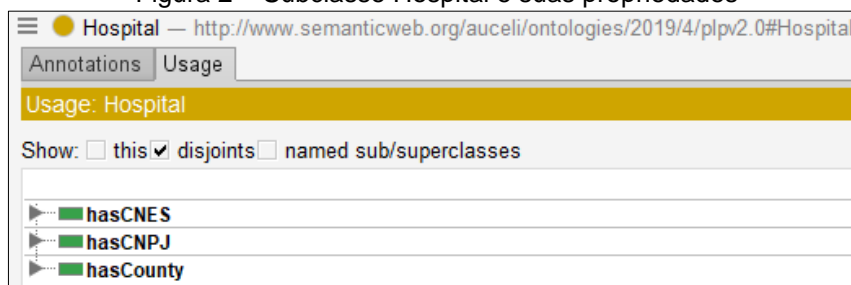
As propriedades de dados foram as que mais sofreram alterações de utilização na versão 2.0 da PLP. Dentre 17 propriedades da ontologia original PLP, 6 delas eram referentes a dados exclusivos da base E-Saude e, portanto, não serão utilizadas no mapeamento da versão 2.0. Entretanto elas não serão excluídas na nova versão da ontologia, possibilitando sua utilização em caso de trabalhos futuros que venham a utilizar a ontologia PLP v2.0 em bases diferentes que contenham tais informações. As Imagens 1 a 3, extraídas do mapeamento realizado na ferramenta Protege, mostram as propriedades referentes a cada classe em sua nova configuração.

Figura 1 – Subclasse Diagnosis e suas propriedades



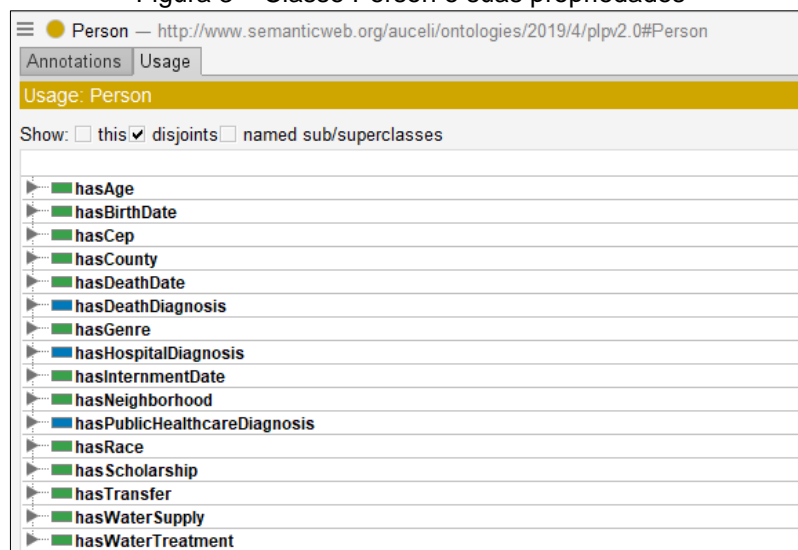
Fonte: Autoria própria (2019).

Figura 2 – Subclasse Hospital e suas propriedades



Fonte: Autoria própria (2019).

Figura 3 – Classe Person e suas propriedades



Fonte: Autoria própria (2019).

Quatro novas propriedades de dados foram incluídas, provenientes de dados complementares da nova base SIASUS-PA, selecionada para o trabalho atual. A Tabela 4 descreve as propriedades de dados da ontologia PLP, as diferenças de utilização e mapeamento de cada versão da ontologia.

Tabela 4 – Ontologia PLP - Propriedades de Dados.

| Propriedades de dados | Descrição | Coluna de mapeamento PLP | Alteração de mapeamento PLP v2.0 |
|-----------------------|---|------------------------------|----------------------------------|
| hasBirthDate | Instância da classe Person, contém a data de nascimento do paciente, mapeável nas três bases. | Data de nascimento - E-Saude | NASC - SIHSUS |

Tabela 4 – Ontologia PLP - Propriedades de Dados.

(continua)

| Propriedades de dados | Descrição | Coluna de mapeamento PLP | Alteração de mapeamento PLP v2.0 |
|------------------------------|--|--|--|
| hasCep | Instância da classe Person, contém o CEP de residência do paciente. | CEP - SIHSUS | |
| hasCID | Instância da classe Diagnosis contém o código de diagnóstico. | Codigo do CID - E-Saude DIAG_PRINC - SIHSUS CAUSABAS - SIM | PA_CIDPRI - SIASUS-PA DIAG_PRINC - SIHSUS CAUSABAS - SIM |
| hasCIDDescription | Instância da classe Diagnosis contém a descrição do diagnóstico. | Descricao do CID - E-Saude | Não será utilizada. |
| hasCNES | Instância da subclasse Hospital, contém o CNES do hospital. | CNES - SIHSUS | |
| hasCNPJ | Instância da subclasse Hospital, contém o CNPJ do hospital. | CGC_HOSP - SIHSUS | |
| hasCounty | Instância que cria um relacionamento entre a classe Person e a subclasse Hospital, contém o município do paciente e do hospital. | Municipio - E-Saude MUNIC_RED - SIHSUS | AP_UFMUN - SIASUS-PA MUNIC_RED - SIHSUS |
| hasDeathDate | Instância da classe Person, contém a data da morte do paciente. | DTOBITO - SIM | |
| hasGenre | Instância da classe Person, contém o gênero do paciente. | Sexo - E-Saude | PA_SEXO - SIASUS-PA |

Tabela 4 – Ontologia PLP - Propriedades de Dados.

(continua)

| Propriedades de dados | Descrição | Coluna de mapeamento PLP | Alteração de mapeamento PLP v2.0 |
|------------------------------|--|-----------------------------------|---|
| hasInternmentDate | Instância da classe Person, contém a data de internamento do paciente. | DT_INTER - SIHSUS | |
| hasNeighborhood | Instância da classe Person, contém o bairro do paciente. | Bairro - E-Saude | Não será utilizada. |
| hasRace | Instância da classe Person, contém a raça do paciente. | RACACOR - SIM | |
| hasScholarship | Instância da classe Person, contém a escolaridade do paciente. | ESC - SIM | |
| hasUpaCode | Instância da classe Upa, contém o código da UPA. | Codigo da Unidade - E-Saude | Não será utilizada. |
| hasUpaName | Instância da classe Upa, contém o nome da UPA. | Descricao da Unidade - E-Saude | Não será utilizada. |
| hasWaterSupply | Instância da classe Person, contém o nível de abastecimento da água na casa do paciente. | Abastecimento - E-Saude | Não será utilizada. |
| hasWaterTreatment | Instância da classe Person, contém o nível de tratamento de água na casa do paciente. | Tratamento no Domicilio - E-Saude | Não será utilizada. |
| hasAge | Instância da classe Person, contém idade do paciente. | | PA_IDADE - SIASUS-PA |
| hasTransfer | Instância da classe Person, indica ocorrência de transferência do paciente. | | PA_TRANSF - SIASUS-PA |

Tabela 4 – Ontologia PLP - Propriedades de Dados.

(conclusão)

| Propriedades de dados | Descrição | Coluna de mapeamento PLP | Alteração de mapeamento PLP v2.0 |
|-----------------------|---|--------------------------|----------------------------------|
| hasProcedureDate | Instância da classe Diagnosis, contém a data do procedimento. | | PA_CMP - SIASUS-PA |
| hasProcedureCode | Instância da classe Diagnosis, contém código de procedimento. | | PA_PROC_ID - SIASUS-PA |

Fonte: Autoria própria (2019).

3.3. PREPARAÇÃO DOS DADOS ABERTOS

Inicialmente a base SIASUS-PA era composta por arquivos obtidos no formato DBC. Foi realizada uma conversão de tipo de arquivo por meio da ferramenta TabWin, disponibilizada pelo próprio *sites* do DATASUS. Os arquivos foram traduzidos para o formato CSV para serem manipulados de forma mais clara, por meio da linguagem Python, na ferramenta Jupyter. Cada mês de referência possui 2 arquivos com dados a nível estadual, compostos por 60 colunas. Os arquivos foram analisados dois a dois, agrupados por mês de referência, somando-se então 24 arquivos no total.

Foram selecionadas as colunas referentes aos dados do trabalho inicial de Auceli e Garcia (2019, no prelo). Para os dados referentes à base SIASUS-PA, inicialmente haviam 60 colunas. Após aplicação de limpeza e filtragem de dados, foram reduzidas a 11 colunas, sendo elas:

- **PA_SEXO**: Sexo do paciente (Compõe o UID);
- **PA_IDADE**: Idade do paciente em anos (Compõe o UID);
- **PA_MUNPCN**: Código do município de residência do paciente (Compõe o UID), fonte IBGE;
- **PA_CNPJCPF**: CNPJ do estabelecimento de atendimento;
- **PA_CODUNI**: Código do CNES do estabelecimento de saúde;
- **PA_CMP**: Data da realização do procedimento (formato AAAAMMDD);
- **PA_PROC_ID**: Código de procedimento ambulatorial;
- **PA_OBITO**: Indicador de óbito (1 para sim, 0 para não);
- **PA_TRANSF**: Indicador de transferência (1 para sim, 0 para não);

- **PA_CIDPRI**: CID principal, Classificação Internacional de Doença;
- **PA_RACACOR**: Identificador raça/cor do paciente (identificadores de 1 a 5).

Na filtragem dos dados referentes à base SIHSUS, foram utilizadas as mesmas colunas inicialmente selecionadas no trabalho de Auceli e Garcia (2019, no prelo). Nessa base, cada mês de referência possuía um arquivo separado. Para a análise, os arquivos foram compilados e, após a limpeza, houve uma redução de 113 colunas para 13 colunas, sendo elas:

- **SEXO**: Sexo do paciente (Compõe o UID);
- **IDADE**: Idade do paciente (Compõe o UID);
- **MUNIC_RES**: Código do município de residência do paciente (compõe o UID), fonte IBGE;
- **NASC**: Data de nascimento do paciente (formato AAAAMMDD);
- **CEP**: CEP do paciente;
- **CGC_HOSP**: CNPJ do estabelecimento;
- **CNES**: Código CNES do hospital;
- **DIAG_PRINC**: Código do diagnóstico principal (CID10);
- **DT_INTER**: Data de internação (formato AAAAMMDD);
- **DT_SAIDA**: Data de saída (formato AAAAMMDD);
- **MORTE**: Indicador óbito (1 para sim, 0 para não);
- **MUNIC_MOV**: Código do município do estabelecimento, fonte IBGE;
- **NASC**: Data de nascimento do paciente (formato AAAAMMDD).

Na filtragem dos dados referentes à base SIM, também foram utilizadas as mesmas colunas inicialmente selecionadas no trabalho de Auceli e Garcia (2019, no prelo). Nessa base há apenas um registro referente ao ano e ao estado (UF) selecionados, que no caso foram os dados do ano de referência de 2017, no escopo da cidade de São Paulo. Após passar por limpeza de dados, onde foram removidas as inconsistências, reduziu-se os dados ao escopo municipal. As colunas foram filtradas e reduzidas de 92 colunas iniciais para 8 colunas, sendo elas:

- **IDADE**: Idade do paciente (compõe o UID)
- **SEXO**: Sexo do paciente (compõe o UID);
- **CODMUNRES**: Código do município de residência do paciente (compõe o UID),

fonte IBGE;

- **DTNASC:** Data de nascimento do paciente (formato AAAAMMDD);
- **CAUSABAS:** Causa básica da morte, conforme a Classificação Internacional de Doença (CID);
- **RACACOR:** Indicação de raça/cor do paciente (identificadores de 1 a 5);
- **ESC:** Escolaridade, anos de estudo concluídos (identificador numérico a partir de 0);
- **DTOBITO:** Data do óbito do paciente (formato AAAAMMDD).

Os resultados de números de registros nas etapas de filtragem de dados foram descritos na Tabela 5.

Tabela 5 – Tabela de filtragem de dados por base utilizada.

| Base | Registros Iniciais | Registros filtrados: estado de SP | Registros filtrados: cidade de São Paulo |
|-------------|---------------------------|--|---|
| SIASUS | 42 milhões | 30.649 | 13.591 |
| SIHSUS | 2 milhões | 243.230 | 58.066 |
| SIM | 2,9 milhões | 293.797 | 75899 |

Fonte: Autoria própria (2019).

Todas as bases de dados selecionadas neste trabalho foram filtradas de acordo com os mesmos parâmetros para eliminar as inconsistências de dados. Cada registro foi convertido em triplas RDF e tendo conexão com outros registros. Havendo inconsistência ou falta de alguma informação, esse vínculo pode não ser realizado e o registro fica inutilizável para o objetivo deste trabalho.

A limpeza nas bases de dados deste trabalho, se deu da seguinte forma, segundo os critérios utilizados por Auceli e Garcia (2019, no prelo):

- **Normalização da informação de sexo do paciente** - na base SIHSUS, o identificador de gênero do paciente era demonstrado com o número “1” para masculino ou “3” para feminino. Os dados que não continham nenhuma das informações foram removidos.
- **Remoção de registros de SIASUS informação negativa para transferência ou óbito;**

- **Normalização de informação de Raça/Cor do paciente** - nas bases SIASUS-PA e SIM. A normalização foi realizada conforme a Tabela 6:

Tabela 6 - Tabela de conversão de campos Raça/Cor.

| Valor Inicial | Conversão |
|---------------|-----------|
| 1 | Branca |
| 2 | Preta |
| 3 | Amarela |
| 4 | Parda |
| 5 | Indígena |

Fonte: Autoria própria (2019).

Ao utilizar os dados da base SIASUS-PA, não foram encontradas informações referentes à “data de nascimento” do paciente, mapeada na ontologia original de Auceli e Garcia (2019, no prelo) pela E-Saude. Essa informação era um dos componentes do identificador único de paciente, composto pela concatenação das colunas: “**Sexo**”, “**Data de Nascimento**” e “**Cidade**”, criado para cada um dos registros nas três bases: o *Unique Identifier* (UID), utilizado para gerar informações referentes aos perfis de pacientes.

Com a ausência de informações correspondentes, foi decidido criar uma nova versão da ontologia de Auceli e Garcia (2019, no prelo), denominada “PLP v2.0”, com inclusão de novas propriedades e a criação de um novo identificador único. Para não gerar inconsistência de dados para usuários que os utilizarem a versão inicial da ontologia, o identificador único para referência na nova versão da ontologia será definido como o indicador “UID2”, conservando assim a configuração inicial da ontologia para possíveis interações como dados filtrados pela ontologia original.

No identificador UID2, anteriormente composto pela concatenação das colunas: “**sexo**”, “**data de nascimento**” e “**cidade**”, o campo com informação de “Data de Nascimento” foi substituído pelo campo “idade” pela ausência de informação comum nas três bases de dados. Para a criação do UID2, foram utilizadas as colunas:

- “**Sexo**” – 1 caractere: “F” ou “M”;
- “**Idade**” – 3 caracteres: de 000 a 199;
- “**Cidade**” – 6 caracteres com o código do município segundo IBGE.

Com a concatenação das colunas indicadas, criou-se o identificador UID2 composto de 10 caracteres, como exemplificado abaixo:

- **SIM:** F054352230;
- **SIHSUS:** M082355030;
- **SIASUS-PA:** M057351910.

Decidiu-se pela permanência de ambos os identificadores UIDs na ontologia PLP v2.0, com base no guia de estudos de boas práticas de Rautenberg *et al.* (2018, p. 44), que diz que o “reúso de ontologias e vocabulários existentes é um fator de inclusão e expansão da *web* de dados, de forma a possibilitar o uso de todos os seus benefícios”. Baseando-se neste guia de boas práticas, foram construídas as Tabelas 02 a 04, na Seção 4.2, contendo a descrição das estruturas da ontologia para cada uma das versões da ontologia PLP, explicando diferenças e fornecendo ao usuário o conhecimento de versionamento para a possibilidade de utilizar dados filtrados pela ontologia original e também pela versão v2.0, para poder conectar ambas as informações geradas.

Foi incluída também uma propriedade de referência na nova ontologia com a informação de versão baseada na linguagem OWL, utilizando a propriedade “owl:versionInfo”, seguindo a lista de “Melhores Práticas” descrita por Rautenberg *et al.* (2018), com ênfase para a Melhor Prática número 7, sobre fornecimento de indicadores de versão:

Tal identificação única permite aos consumidores de dados determinarem como os dados têm mudado ao longo do tempo e, especificamente qual versão de um conjunto está se considerando. (RAUTENBERG *et al.*, 2018, p. 38)

Após a modelagem dos dados extraídos das bases, foram criadas 3 tabelas em banco de dados relacional PostgreSQL, cada uma referente a uma base de dados: “siasus”, “sihsus” e “sim”. Com os dados alocados no banco de dados foi possível realizar o mapeamento das colunas, transformando-as em classes, propriedades de objetos e dados, por meio do *framework* Ontop dentro da ferramenta Protege.

Nas Tabelas 7, 8 e 9 são apresentados os mapeamentos realizados para as classes e propriedades: “Person”, “Diagnosis” e “Hospital”, consecutivamente, assim

como a consulta SQL referente ao mapeamento, destacando as principais alterações no mapeamento das triplas em cada classe.

Tabela 7 – Tabela de mapeamento de triplas: Classe Person

| Mapeamento de triplas: classe Person | Query SQL |
|---|---|
| :{uid2} a :Person ; :hasAge {idade} ; :hasBirthDate {datanasc} ; :hasCep {cep} ; :hasScholarship {escolaridade} ; :hasGenre {sexo} ; :hasCounty {municipioresidencia} ; :hasRace {racacor} ; :hasInternmentDate {datainternamento} ; :hasTransfer {transfPac} ; :hasProcedureDate {dataProcedimento} ; :hasProcedureCode {codProcedimento} ; :hasDeathDiagnosis :{cidbasico} ; :hasHospitalDiagnosis :{codigocid} ; :hasPublicHealthcareDiagnosis :{cidprinc} . | SELECT a.uid2, a.idade, h.datanasc, codigocid, cidbasico, cidprinc, h.cep, m.escolaridade, a.sexo, h.datainternamento, h.municipioresidencia, m.racacor, a.transfPac, a.dataProcedimento, a.codProcedimento FROM siasus a,sihsus h, sim m WHERE (a.uid2=h.uid2 AND h.uid2=m.uid2) |

Fonte: Autoria própria (2019).

Tabela 8 – Tabela de mapeamento de triplas: Classe Diagnosis

| Mapeamento de triplas: classe Diagnosis | Query SQL |
|---|-------------------------------------|
| :{cidPrinc} a :Diagnosis ; :hasCID {cidPrinc} . | SELECT cidPrinc FROM siasus |
| :{codigocid} a :Diagnosis ; :hasCID {codigocid} . | SELECT codigocid FROM sihsus |
| :{cidbasico} a :Diagnosis ; :hasCID {cidbasico} . | SELECT cidbasico FROM sim |

Fonte: Autoria própria (2019).

Tabela 9 – Tabela de mapeamento de triplas: Classe Hospital

| Mapeamento de triplas: classe Hospital | Query SQL |
|--|--|
| :{cnes} a :Hospital ; :hasCNPJ {cnpj} ; :hasCNES {cnes} ; :hasCounty {municipioestabelecimento} . | SELECT cnpj, cnes, municipioestabelecimento FROM sihsus |

Fonte: Autoria própria (2019).

4. RESULTADOS

Na tarefa de reúso da ontologia PLP, a primeira barreira clara de adaptação foi a interligação de informações das novas bases de dados selecionadas para o estudo. Apenas a mudança de bases ou parâmetros de tipos de dados, gera vários impactos na reutilização de uma ontologia, pela necessidade de novas abordagens na seleção e modelagem de dados, até que se possa chegar à etapa de efetivamente trabalhar com própria ontologia.

Rautenberg *et al.* (2018), em seu guia para publicação de dados conectados, apresenta algumas “Melhores Práticas” estabelecidas no Consórcio W3C⁴ para consumo e publicação de dados na *web*. Sob essas “Melhores Práticas” (MPs), são identificados pelos autores 8 benefícios relacionados às práticas, descritas na Seção 4.1, juntamente com a sua utilização para tomadas de decisões neste trabalho.

4.1. RESULTADO E DISCUSSÃO SOBRE O REÚSO

No desenvolvimento desse trabalho, em cada etapa de aplicação da ontologia, ou de preparação dos dados para o estudo da ontologia, houve momentos de dificuldades quando houve a necessidade de tomada de decisão sobre a abordagem a se utilizar, para que fossem alcançadas as reais vantagens do reúso da ontologia PLP.

Para que se mantivesse a integridade da ontologia, mesmo frente a necessidades de adaptações, utilizou-se o guia prático de Rautenberg *et al.* (2018), que se baseia em “Melhores Práticas” para obter o melhor resultado na utilização de dados abertos.

Na sequência serão descritas quais foram as vantagens, dificuldades e decisões na construção do trabalho atual, como uma nova versão da ontologia PLP. Foram tomadas precauções para não tirar o foco do trabalho original dos autores Auceli e Garcia (2019, no prelo), mantendo assim a integridade da ontologia PLP em cada ângulo dos 8 benefícios estipulados por Rautenberg *et al.* (2018), na utilização das “Melhores Práticas” de W3C, os quais são:

⁴ <https://www.w3.org/TR/dwbp/>

1. **Compreensão:**

- a. Definições de informação de alterações da nova versão da ontologia PLP. Foram criadas tabelas de comparação da estrutura inicial da ontologia PLP com as alterações que surgiram na PLP v2.0;
- b. Foi mantida a indicação da origem e formatação dos dados, como por exemplo: formato de data, sequência da concatenação de UID, indicadores de sexo e raça, informação da fonte de códigos obtidos de instituições específicas (CNJP, CNES, código de município);

2. **Processabilidade**

- a. Este trabalho propõe a disponibilização de dados, tanto sobre os dados formatados quanto de análise, disponibilizando em vários formatos, assim como foram obtidos e convertidos (DBC, DBF, CVS), para que outras aplicações possam ter fácil acesso sem a necessidade de retrabalho na filtragem de dados.
- b. O enriquecimento dos dados, poderia ocorrer com a construção de uma API, onde se concentrariam os dados utilizados em seus formatos originais e suas segmentações, assim como as versões da ontologia.

3. **Descoberta** – É importante, ao final de um trabalho, disponibilizar informações contendo: o arquivo final da ontologia, os dados utilizados e os resultados. Juntamente com essas informações, é importante documentar as características dos dados utilizados, com definições e formatação dos dados. Para a base SIASUS-PA que trazia novos tipos de dados para a ontologia, houve o trabalho de encontrar sua documentação, foi então adicionado o Anexo A, neste trabalho, concentrando a informação já adquirida sobre os dados, que será disponibilizada juntamente com publicação da ontologia.

4. **Reúso** – A reutilização de colunas sobre as bases SIM e SIHSUS, que já haviam sido mapeadas para a ontologia PLP original, facilitou o trabalho de identificação de colunas que poderiam ser selecionadas em referência ao escopo deste trabalho e aos novos tipos de dados provenientes da base SIASUS.

5. Confiança

- a. Este trabalho dedicou atenção ao esclarecimento das colunas de dados e propriedades da ontologia que não seriam utilizados na nova versão da ontologia PLP, v2.0. Essa ação se deve à preocupação de manter o usuário informado sobre quais dados ele terá disponíveis para sua modelagem e análise, para que não haja quebras inesperadas de informações.
- b. Todas as alterações ocorridas na construção da versão 2.0 da ontologia PLP, serão informadas aos autores da ontologia original, no final do trabalho. É importante também realizar um *feedback* sobre melhorias ou ajustes na reutilização do dado, para que os autores possam continuar seus trabalhos ou até corrigir informações e dados, caso necessário.
- c. É importante seguir as licenças de publicação do autor da informação. Por mais que esse tipo de especificação não estivesse presente no trabalho de Auceli e Garcia (2019, no prelo), foi determinado desde o início deste trabalho que se trataria de um réuso da ontologia. Mantendo assim a autoria e direitos da ontologia ainda no nome dos autores da ontologia PLP.

6. Ligação – Uma dificuldade encontrada na realização deste trabalho foi não haver disponibilização dos dados utilizados no trabalho de Auceli e Garcia (2019, no prelo). Pois, caso houvesse a disponibilização do conjunto ou subconjuntos de dados utilizados no trabalho, poderia ser feito um estudo de ligação entre dados em versões diferentes da mesma ontologia. O que poderia trazer novos conhecimentos quanto às práticas de reutilização de ontologias.

7. Acesso – Assim como já mencionado anteriormente, é proposta neste trabalho a publicação dos dados utilizados na construção da nova versão da ontologia para que uma futura reutilização de qualquer uma das versões da ontologia PLP seja facilitada pelo acesso aos conjuntos de dados modelados neste trabalho. Juntamente, será disponibilizada a documentação dos dados, sendo também este estudo uma fonte de

informação de estruturas, do método e etapas da reutilização da ontologia PLP.

- 8. Interoperabilidade** – Um facilitador neste trabalho foi a utilização da semântica formal por meio de vocabulários simples. Esse e outros facilitadores deste trabalho devem ser reportados aos autores da ontologia PLP, para que vejam os resultados e os pontos positivos de seu trabalho e valorizem a construção prática da ontologia, que facilita o reestudo.

A Tabela 10 aponta os benefícios citados por Rautenberg *et al.* (2018) decorrentes da adoção das Melhores Práticas da Convenção W3C. A descrição individual de cada Melhor Prática e sua observação direta pode ser vista na Tabela 11.

Tabela 10 – Tabela de benefícios decorrente da adoção das Melhores Práticas

| Benefício | Melhores Práticas adotadas |
|--------------------|--|
| Compreensão | MP1, MP13, MP16, MP31. |
| Processabilidade | MP1, MP14, MP18, MP23, MP31. |
| Descoberta | MP1. |
| Reúso | MP1, MP13, MP14, MP16, MP18, MP19, MP21, MP22, MP23, MP31, MP33, MP35. |
| Confiança | MP22, MP31, MP33. |
| Ligação | MP18. |
| Acesso | MP18, MP19, MP21, MP22, MP23. |
| Interoperabilidade | MP16, MP23, MP33. |

Fonte: Aatoria própria (2019).

Tabela 11 – Tabela de definições de Melhores Práticas adotadas neste estudo (continua)

| MPs adotadas | Descrição | Abordagem correspondente neste trabalho |
|---------------------|---|--|
| MP1 | Forneça metadados - Deve haver o fornecimento de metadados tanto para humanos como para máquinas de legível por ambos. | APLICADA: Neste trabalho foram construídas tabelas de descrição de dados utilizados e sobre alterações em propriedades da ontologia. Fornecendo metadados com foco na interpretação humana. |

Tabela 11 – Tabela de definições de Melhores Práticas adotadas neste estudo (continua)

| MPs adotadas | Descrição | Abordagem correspondente neste trabalho |
|---------------------|------------------|--|
|---------------------|------------------|--|

| | | |
|------|---|---|
| MP13 | Use representações de dados neutros de localidade - Dados legíveis por máquinas que não são específicos de um idioma ou Cultura em particular são mais duráveis e geram menos interpretações erradas que dados do que dados que utilizam nada sob uma representação específica. Por exemplo dados sobre datas moedas e números podem ter significados diferentes em locais diferentes. | APLICADA: Neste trabalho houve o cuidado de indicar as informações de fontes de referência de dados, assim como indicação de formatação específica sobre data, códigos ou identificadores específicos. |
| MP14 | Forneça dados em múltiplos formatos - Fornecer dados em mais de um formato reduz custo envolvido na transformação dos dados, deve-se considerar a implementação de dados mais propícias ao reuso fornecendo alternativas para que os dados possam ser úteis no futuro. | OBSERVAÇÃO: Neste trabalho observou-se a necessidade de, juntamente com a ontologia PLP, obter-se os dados utilizados em formatos diversos que possam ser reutilizados em qualquer etapa da reaplicação de uma ontologia, ou na construção de novas versões, como ocorreu neste trabalho. |
| MP16 | Escolha o nível de formalização certo - A semântica formal ajuda a estabelecer as especificações que transmitem um significado preciso e detalhado, deve-se verificar a existência de vocabulários padronizados que correspondam às atuais necessidades do trabalho. Recomenda-se o uso de vocabulários mais simples quando forem suficientes para cobrir a semântica. | APLICADA: O nível de simplicidade do vocabulário foi fator essencial neste estudo, para que não se tornasse complexa a análise do vocabulário na identificação e referência dos dados correspondentes nas novas bases de dados utilizadas. |
| MP18 | Forneça subconjuntos para conjuntos volumosos de dados - Conjuntos volumosos de dados são difíceis de armazenar e transferir. Caso os usuários precisem deve ser fornecido um meio para uma recuperação parcial, pois o desempenho de processamento de aplicações é mais eficiente utilizando subconjuntos de dados. | OBSERVAÇÃO: Assim como a observação sobre a M14, a falta da disponibilização dos dados iniciais do trabalho da ontologia PLP original, nos formatos iniciais ou pré-processados foi sentida, não sendo possível sua utilização para comprovação de sucesso no estudo de construção de nova versão da ontologia, como ocorreu neste trabalho. |
| MP19 | Use negociação de conteúdo para fornecer informações de dados em múltiplos formatos - Um conjunto de dados podem ter várias representações em vários tipos de formatos, essas representações podem ser disponibilizadas por meio de uma API a partir de um mesmo URL. | OBSERVAÇÃO: Esta foi uma observação realizada na seção de CONCLUSÃO e uma indicação para trabalhos futuros, a construção de uma API onde se possa disponibilizar dos dados aplicados, assim como informações sobre versões de ontologia, conectadas e referenciadas em um só local. |
| MP21 | Forneça dados atualizados - Disponibilização de dados na <i>web</i> deve corresponder com a sua data de criação coleta ou processamento. Pode-se criar uma programação regular para envio de novas versões do conjunto de dados para <i>web</i> para evitar que os dados se tornem desatualizados. | APLICADA: No mesmo local onde será disponibilizada a ontologia de Auceli e Garcia (no prelo), será documentado o estudo de reaplicação e disponibilizada a nova versão da ontologia, para manter as versões sempre atualizadas referentes às suas datas de publicações. |

Tabela 11 – Tabela de definições de Melhores Práticas adotadas neste estudo (conclusão)

| MPs adotadas | Descrição | Abordagem correspondente neste trabalho |
|--------------|-----------|---|
|--------------|-----------|---|

| | | |
|------|---|---|
| MP23 | Torne os dados disponíveis utilizando uma API – Uma API oferece mais flexibilidade capacidade de processamento para consumidores de dados uma permite uso de dados em tempo real, filtragem em requisições e a capacidade de trabalhar com dados em nível atômico. | OBSERVAÇÃO: Complementando a observação da MP21, a construção de uma API possibilita a atualização em tempo real a concentração de vários formatos de dados, sendo necessário apenas parâmetros de chamada para que se obtenha os dados mais recentes sempre na mesma chamada de requisição. |
| MP31 | Enriqueça dados gerando novos dados - O enriquecimento dos dados pode melhorar seu processamento principalmente em dados não estruturados, pode ocorrer ao completar-se valores não preenchidos, adicionando novos atributos e medidas. | APLICADA: O enriquecimento de dados neste trabalho se deu pela criação na nova versão 2.0 da ontologia PLP, realizando adição de propriedades que enriquecem a análise do trabalho e a informação final. |
| MP33 | Forneça feedback para o publicador original - Relatar o uso de dados um publicador ajuda justificar seus esforços para a divulgação dos dados, além de auxiliar diretamente na melhoria do conjunto de dados para usuários futuros. | APLICADA: Assim como durante a realização deste trabalho ocorreu contato constante com os autores originais da ontologia, ao finalizar o estudo e reaplicação da ontologia PLP, os autores receberão o estudo assim como <i>feedback</i> e anotações de resultados do trabalho. |
| MP35 | Cite o publicador original - identificação da fonte dos dados é um dos principais indicadores de confiabilidade, além de que citar os publicadores de dados ajuda acreditar seu trabalho representando reconhecimento a eles incentivando a continuar o compartilhamento de dados. | APLICADA: Assim como citado anteriormente nesta seção, na observação do benefício 5: Confiança, foi determinado desde o início que este trabalho se trataria de um reuso da ontologia, mantendo assim a autoria e direitos da ontologia ainda no nome dos autores da ontologia PLP. |

Fonte: Autoria própria (2019).

Nesse trabalho, foi necessária a criação de uma nova versão da ontologia, por haver a necessidade de uma nova formatação dos identificadores únicos para o perfil do paciente. Porém acredita-se que as dificuldades encontradas durante este trabalho não têm grandeza suficiente a ponto de desestimular a continuidade do estudo de reaplicação da ontologia PLP em bases de dados diferentes. Há uma atenção que se necessita ter na instrução da seleção e filtragem das bases de dados, para que possa haver a disponibilidade e reaproveitamento também de dados já analisados e utilizados anteriormente.

5. CONCLUSÃO

Por meio da pesquisa bibliográfica, foi possível identificar a importância de dados que sejam disponibilizados em bases de dados abertos, permitindo assim o acesso facilitado a qualquer pessoa ou instituição que possa reutilizar estudos já aplicados,

poupando novos usuários do retrabalho e até mesmo esforços financeiros para a análise de domínios já estudados.

Na pesquisa bibliográfica, foi possível obter o conhecimento essencial para o projeto, no que diz respeito à importância de uma ontologia ser descritivamente ampla o suficiente para que possa ser aplicada em diversas bases do mesmo domínio, e ainda permitir a adição de novas definições que a enriqueçam. Destaque deve ser dado ao guia pelas Melhores Práticas de publicação de Dados Abertos Conectados apresentado por Rautenberg *et al.* (2018), que auxilia nas decisões de abordagens quando se trabalha com produção de dados conectados.

Acredita-se que esse trabalho realizou um bom reaproveitamento da ontologia PLP dos autores Auceli e Garcia (2019, no prelo), encontrando algumas dificuldades na adaptação no uso de novos dados de diferentes origens, porém realizar a seleção e filtragem dos dados foi o que trouxe mais desafios neste trabalho. O mapeamento da ontologia, com pequenas adaptações incluindo propriedades de objetos foi completamente intuitivo e facilitado pelos registros de construção deixados pelos autores da ontologia PLP.

Foram identificadas, durante o estudo de melhores práticas, algumas ações facilitadoras à reaplicação de uma ontologia que podem ser realizados como trabalhos futuros. Até mesmo este estudo poderia resultar em diferentes resultados e análises mais avançadas, caso essas ações tivessem sido apontadas durante o trabalho de Auceli e Garcia (2019, no prelo).

A primeira ação facilitadora identificada, está ligada a Melhor Prática 19 citada por Rautenberg *et al.* (2018), referente ao fornecimento de conjuntos de dados em múltiplos formatos de arquivos. Pois se dados de entrada de um estudo são transformados e filtrados para permanecer com apenas as informações necessárias, a sugestão é armazenar os arquivos iniciais, juntamente com os dados pré-processados, antes de haver cortes e filtragem. Isso possibilita a reutilização da mesma fonte de dados, sem a necessidade do processamento inicial de normalização de dados. Com dados de diversos formatos, cada um pode ser utilizado por um tipo de estudo.

Isto remete à segunda ação facilitadora, sobre tornar os dados disponíveis em uma API, citado pelos mesmos autores da ação anterior, se baseia nas Melhores Práticas 21 e 23. Para manter os dados de uma linha de estudo atualizados, é

importante concentrar os dados e arquivos de análises, assim como versões da ontologia, sobre o mesmo local de armazenamento. Sendo uma API uma forma de fazer uma requisição e obter dados armazenados sobre um domínio específico, fica neste trabalho a sugestão para que futuros estudos da área envolvam a construção de uma API que permita o acesso e obtenção de dados de armazenamento de versionamentos de ontologias, podendo incluir vocabulários e outras estruturas de *Linked Data*. Dessa forma, cria-se um meio facilitado para que haja um maior foco dos estudos específicos, sem que necessitem realizar vários retrabalhos antes de chegar ao seu objeto de estudo.

A comparação das versões da ontologia apresentou modelos bem semelhantes e existência da possibilidade de conectar dados utilizados de versões diferentes, porém com o conhecimento prévio, para que haja a construção de um indicador comum. O reúso de ontologias precisa ser mais difundido e praticado. A maior contribuição deste trabalho, por suas etapas e análises finais, foi demonstrar que a reutilização é possível e custa menos esforços do que um desenvolvimento e construção de algo novo e, desnecessariamente, específico.

REFERÊNCIAS

- ARAÚJO, L. R.; SOUZA, J. F. de **Aumentando a transparência do governo por meio da transformação de dados governamentais abertos em dados ligados**. Revista Eletrônica de Sistemas de Informação, v. 10, n. 1, 2011. Disponível em <<http://www.periodicosibepes.org.br/index.php/reinfo/article/view/880/pdf>> Acesso em 05 ago. 2019
- AUCELI, P. H. S., GARCIA, Y. A. **Integração semântica de bases de dados abertos de saúde de Curitiba**: um estudo de caso com atendimentos com desfecho em óbito. Trabalho de Conclusão de Curso (Graduação em Bacharelado em Sistemas de Informação) – Universidade Tecnológica Federal do Paraná, Curitiba, 2019 (no prelo).
- BIZER, C.; HEATH, T.; BERNERS-LEE, T. **Linked data: The story so far**. (2011) In: Semantic services, interoperability and web applications: emerging concepts. IGI Global, 2011. p. 205-227.
- BONTAS, E. P.; MOCHOL, M.; TOLKSDORF, R. **Case studies on ontology reuse**. In: Proceedings of the IKNOW05 International Conference on Knowledge Management. 2005. p. 345.
- DA ROCHA, R. P. **Fabrico/Ciência: Um ambiente Linked Data para o mapeamento da ciência**. Em Questão, v. 18, n. 3, p. 281-297, 2012. Disponível em <<https://dialnet.unirioja.es/servlet/articulo?codigo=6067111>>. Acesso em: 06 ago. 2019
- DORAN, P.; TAMMA, V.; IANNONE, L. **Ontology module extraction for ontology reuse**: an ontology engineering perspective. In: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management. ACM. 2007. p. 61-70.
- GUARINO, N.; OBERLE, D.; STAAB, S. **What is an ontology?** In: Handbook on ontologies. Springer, Berlin, Heidelberg, 2009. p. 1-17.
- ISOTANI, S., BITTENCOURT, I. I. **Dados Abertos Conectados**: Em Busca da Web do Conhecimento. São Paulo: Novatec Editora. 2015. p.176.
- MARCONDES, C. H. **“Linked data” – dados interligados - e interoperabilidade entre arquivos, bibliotecas e museus na web**. Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 17, n. 34, p. 171-192, ago. 2012.
- MARTINS, A. F. **Construção de Ontologias de Tarefa e sua Reutilização na Engenharia de Requisitos**. Dissertação de Mestrado. Universidade Federal do Espírito Santo. 2009. Disponível em <<http://repositorio.ufes.br/jspui/handle/10/6371>>. Acesso em: 8 ago. 2019

MCGUINNESS, D. L. *et al.* **OWL web ontology language overview**. W3C recommendation, 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-owl-features-20040210/>>. Acesso em 02 out. 2019.

PAN, J. Z.; SERAFINI, L.; ZHAO, Y. **Semantic import**: An approach for partial ontology reuse. In: Proceedings of the 1st International Conference on Modular Ontologies-Volume 232. CEUR-WS. 2006. p. 71-84.

Portal da saúde SUS: Transferência/download de arquivos. Disponível em <<http://www2.datasus.gov.br/DATASUS/index.php?area=0901>>. Acessado em: 19 set. 2019. 12, 17

RAUTENBERG, S. *et al.* **Guia Prático para Publicação de Dados Abertos Conectados na Web**. Curitiba: Appris Editora e Livraria Eireli-ME, 2019.

SCHIESSL, J. M. **Ontologia**: o termo e a ideia. Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, v. 12, n. 24, p. 172-181, 2007.

SIQUEIRA, S. *et al.* **Information Systems based on (Linked) Open Data: From Openness to Innovation**. I GranDSI-BR-Grand Research Challenges in Information Systems in Brazil, v. 2026, p. 52-61, 2016. Disponível em <http://www2.sbc.org.br/ce-si/arquivos/GranDSI-BR_Ebook-Final.pdf#page=52>. Acesso em: 06 ago. 2019

ZAIDAN, F. H.; BAX, M. P. **Linked Open Data como forma de agregar valor às informações clínicas**. AtoZ: novas práticas em informação e conhecimento, v. 2, n. 1, p. 44-59, 2013. Disponível em: <<http://www.atoz.ufpr.br>>. Acesso em: 01 ago. 2019.

World Wide Web Consortium. **Best practices for publishing linked data**. 2014.

ANEXOS

**ANEXO A - LAYOUT DO ARQUIVO DE PRODUÇÃO AMBULATORIAL:
PAUFAAMM.DBF**

(continua)

| SEQ | CAMPO | TIPO E TAM | DESCRIÇÃO |
|-----|------------|------------|---|
| 1 | PA_CODUNI | CHAR (7) | Código do Estabelecimento no CNES (Cadastro Nacional de Estabelecimentos de Saúde) ⁶ |
| 2 | PA_GESTAO | CHAR (6) | Código da Unidade da Federação ⁷ (IBGE) + Código do Município (IBGE) do Gestor, ou UF0000 se o estabelecimento estiver sob Gestão Estadual |
| 3 | PA_CONDIC | CHAR (2) | Sigla do Tipo de Gestão no qual o Estado ou Município está habilitado |
| 4 | PA_UFMUN | CHAR (6) | Unidade da Federação + Código do Município onde está localizado o estabelecimento |
| 5 | PA_REGCT | CHAR (4) | Código da Regra Contratual |
| 6 | PA_INCOUT | CHAR (4) | Incremento Outros |
| 7 | PA_INCURG | CHAR (4) | Incremento Urgência |
| 8 | PA_TPUPS | CHAR (2) | Tipo de Estabelecimento |
| 9 | PA_TIPPRE | CHAR (2) | Tipo de Prestador |
| 10 | PA_MN_IND | CHAR (1) | Estabelecimento Mantido / Individual |
| 11 | PA_CNPJCPF | CHAR (14) | CNPJ do Estabelecimento executante |
| 12 | PA_CNPJMNT | CHAR (14) | CNPJ da Mantenedora do estabelecimento ou zeros, caso não a tenha |
| 13 | PA_CNPJ_CC | CHAR (14) | CNPJ do Órgão que recebeu pela produção por cessão de crédito ou zeros, caso não o tenha |
| 14 | PA_MVM | CHAR (6) | Data de Processamento / Movimento (AAAAMM) |
| 15 | PA_CMP | CHAR (6) | Data da Realização do Procedimento / Competência (AAAAMM) |
| 16 | PA_PROC_ID | CHAR (10) | Código do Procedimento Ambulatorial |
| 17 | PA_TPFIN | CHAR (2) | Tipo de Financiamento da produção |
| 18 | PA_SUBFIN | CHAR (4) | Subtipo de Financiamento da produção |
| 19 | PA_NIVCPL | CHAR (1) | Complexidade do Procedimento |
| 20 | PA_DOCORIG | CHAR (1) | Instrumento de Registro (conforme explicado na página 2) |
| 21 | PA_AUTORIZ | CHAR (13) | Número da APAC ou número de autorização do BPA-I, conforme o caso. No BPA-I, não é obrigatório, portanto, não é criticado. Lei de formação: UFAATsssssssd, onde: UF – Unid. da Federação, AA – ano, T – tipo, ssssss – sequencial, d – dígito |
| 22 | PA_CNSMED | CHAR (15) | Número do CNS (Cartão Nacional de Saúde) do profissional de saúde executante |
| 23 | PA_CBOCOD | CHAR (6) | Código da Ocupação do profissional na Classificação Brasileira de Ocupações ⁸ (Ministério do Trabalho) |
| 24 | PA_MOTSAI | CHAR (2) | Motivo de saída ou zeros, caso não tenha |
| 25 | PA_OBITO | CHAR (1) | Indicador de Óbito (APAC) |
| 26 | PA_ENCERR | CHAR (1) | Indicador de Encerramento (APAC) |
| 27 | PA_PERMAN | CHAR (1) | Indicador de Permanência (APAC) |

(continua)

| SEQ | CAMPO | TIPO E TAM | DESCRIÇÃO |
|-----|------------|----------------|--|
| 28 | PA_ALTA | CHAR (1) | Indicador de Alta (APAC) |
| 29 | PA_TRANSF | CHAR (1) | Indicador de Transferência (APAC) |
| 30 | PA_CIDPRI | CHAR (4) | CID9 Principal (APAC ou BPA-I) |
| 31 | PA_CIDSEC | CHAR (4) | CID Secundário (APAC) |
| 32 | PA_CIDCAS | CHAR (4) | CID Causas Associadas (APAC) |
| 33 | PA_CATEND | CHAR (2) | Caráter de Atendimento (APAC ou BPA-I) |
| 34 | PA_IDADE | CHAR (3) | Idade do paciente em anos |
| 35 | PA_FLIDADE | CHAR (3) | Idade mínima do paciente para realização do procedimento |
| 36 | IDADEMIN | CHAR (3) | Idade máxima do paciente para realização do procedimento |
| 37 | IDADEMAX | CHAR (1) | Compatibilidade com a faixa de idade do procedimento (SIGTAP – Sistema de Gerenciamento da Tabela de Procedimentos do SUS): 0 = Idade não exigida; 1 = Idade compatível com o SIGTAP; 2 = Idade fora da faixa do SIGTAP; 3 = Idade inexistente; 4 = Idade EM BRANCO |
| 38 | PA_SEXO | CHAR (1) | Sexo do paciente |
| 39 | PA_RACACOR | CHAR (2) | Raça/Cor do paciente: 01 - Branca, 02 - Preta, 03 - Parda, 04 - Amarela, 05 - Indígena, 99 - Sem informação |
| 40 | PA_MUNPCN | CHAR (6) | Código da Unidade da Federação + Código do Município de residência do paciente ou do estabelecimento, caso não se tenha a identificação do paciente, o que ocorre no (BPA) |
| 41 | PA_QTDPRO | NUMERIC (11) | Quantidade Produzida (APRESENTADA) |
| 42 | PA_QTDAPR | NUMERIC (11) | Quantidade Aprovada do procedimento |
| 43 | PA_VALPRO | NUMERIC (20,2) | Valor Produzido (APRESENTADO) |
| 44 | PA_VALAPR | NUMERIC (20,2) | Valor Aprovado do procedimento |
| 45 | PA_UFDIF | CHAR (1) | Indica se a UF de residência do paciente é diferente da UF de localização do estabelecimento: 0 = mesma UF; 1 = UF diferente |
| 46 | PA_MNDIF | CHAR (1) | Indica se o município de residência do paciente é diferente do município de localização do estabelecimento: 0 = mesmo município; 1 = município diferente |
| 47 | PA_DIF_VAL | NUMERIC (20,2) | Diferença do Valor Unitário do procedimento praticado na Tabela Unificada com Valor Unitário praticado pelo Gestor da Produção, multiplicado pela Quantidade Aprovada |
| 48 | NU_VPA_TOT | NUMERIC (20,2) | Valor Unitário do Procedimento da Tabela VPA |
| 49 | NU_PA_TOT | NUMERIC (20,2) | Valor Unitário do Procedimento da Tabela SIGTAP |
| 50 | PA_INDICA | CHAR (1) | Indicativo de situação da produção produzida: 0 = não aprovado; 5 = aprovado total; 6 = aprovado parcial |

(conclusão)

| SEQ | CAMPO | TIPO E TAM | DESCRIÇÃO |
|------------|--------------|-------------------|--|
| 51 | PA_CODOCO | CHAR (1) | Código de Ocorrência |
| 52 | PA_FLQT | CHAR (1) | Indicador de erro de Quantidade Produzida |
| 53 | PA_FLER | CHAR (1) | Indicador de erro de corpo da APAC |
| 54 | PA_ETNIA | CHAR (4) | Etnia do paciente |
| 55 | PA_VL_CF | NUMERIC (20,2) | Valor do Complemento Federal |
| 56 | PA_VL_CL | NUMERIC (20,2) | Valor do Complemento Local |
| 57 | PA_VL_INC | NUMERIC (20,2) | Valor do Incremento |
| 58 | PA_SRC_C | CHAR(6) | Código do Serviço Especializado / Classificação CBO (de acordo com o CNES) |
| 59 | PA_INE | CHAR(10), | Código de Identificação Nacional de Equipes ¹⁰ , para registrar a atuação das equipes na execução de ações de saúde |
| 60 | PA_NAT_JUR | CHAR(4) | Código da Natureza Jurídica |