UMA ABORDAGEM PARA A PUBLICAÇÃO DE DADOS LIGADOS OBTIDOS A PARTIR DE BASES DE DADOS RELACIONAIS

Clayton Martins Pereira
clayton.martins@inpe.br

José Maria Parente de Oliveira
parente@ita.br

Resumo
Este trabalho tem como objetivo apresentar uma proposta de abordagem para facilitar e automatizar a publicação, na *Web* Semântica, de dados abertos ligados obtidos a partir de bases de dados relacionais (BDRs), por meio da integração entre as diversas ferramentas de software aplicadas neste processo. Oferece ainda uma nova ferramenta que possibilita a customização, de forma semi-automática, do arquivo de mapeamento entre a BDR e o modelo de dados *RDF*, a fim de incorporar a este uma ontologia de domínio fornecida pelo usuário. A abordagem proposta, chamada de *RDB2LOD*, apresenta como diferencial a automatização, por meio de um aplicativo (interface gráfica), das ferramentas utilizadas para a geração e customização deste arquivo de mapeamento, e para a visualização e consulta dos dados abertos ligados obtidos a partir dele. Conclui-se que a abordagem *RDB2LOD* permitirá a publicação na *Web*, em grande escala, de dados ligados obtidos a partir de bases de dados relacionais, de forma a possibilitar amplo acesso a muitos usuários, o que hoje ocorre ainda de forma muito tímida por meio de algumas iniciativas.

Palavras-chave: Dados Ligados. *Web* Semântica. Ontologias. Bases de Dados Relacionais.

Abstract
This work aims to present a proposal of a linked open data approach to facilitate and automate publishing, in the Semantic Web, obtained from relational database (RDB), through the integration between the various software tools used in this process. It also offers a new tool that enables the semi-automatic customization of the mapping among RDB and the RDF data model in order to incorporate a domain ontology supplied by the user. The proposed approach, called RDB2LOD, presents as differential the automation through an application (GUI) of the tools used for generating and customizing of this mapping file, and to view and query the linked open data obtained from it. It is concluded that the RDB2LOD approach will enable web publishing of linked data from relational databases, so as to allow well as a broad access to many users, which still occurs today in a very shy way through some initiatives.

Key Words: Linked data. Semantic web. Ontology. Relational databases.

1 Introdução

Atualmente, há uma vasta quantidade de dados armazenados em Bases de Dados Relacionais (BDRs) (SAHOO et al., 2009).Este é o modelo mais comumente utilizado e constitui o núcleo da maioria dos sistemas de tecnologia da informação (TI) hoje em uso (CHOI et al., 2010).Isso se deve à maturidade e eficiência de sua forma de armazenagem e consulta a dados, além de sua alta confiabilidade e escalabilidade (LING; ZHOU, 2010). Estima-se que 70% dos sítios *Web* são alimentados por dados mantidos em BDRs(SEQUEDA; ARENAS; MIRANKER, 2012).

Por outro lado, busca-se uma evolução da atual "*Web* de documentos" para uma "*Web* de dados",por meio de uma "*Web* Semântica",na qual as informações possam ser disponibilizadas em formato aberto (bruto) e apresentem significados bem definidos, ou seja, exatamente de acordo com o contexto ou domínio de conhecimento ao qual estas sejam aplicadas (semântica), de forma a possibilitar que computadores sejam capazes de processá-las e entendê-las automaticamente (BERNERS-LEE; HENDLER; LASSILA, 2001). Isso permitirá que tais informações sejam compartilhadas a partir de diferentes fontes, além de habilitar seu uso em diferentes contextos por diversas aplicações e agentes de software (LV; MA, 2008). Desta forma, o sucesso da *Web* Semântica depende da criação em massa de dados em formato aberto (LING; ZHOU, 2010).

Neste sentido, disponibilizar as grandes quantidades de dados armazenados em BDRs em formato aberto, por meio da*Web* Semântica,pode permitir que diferentes bases sejam unificadas, bem como seus dados sejam associados a outros dados estáticos ou outras fontes de dados (dados ligados), de forma a se obter um "ganho de informação". Além disso, promoveria ainda a interoperabilidade entre os sistemas de informação existentes e a formação de uma "rede de dados interpretáveis universalmente", a qual poderia ser usada como meio para a construção de aplicações orientadas a dados (LV; MA, 2008).

Diversas ferramentas de *software* têm sido criadas com este propósito, porém verifica-se uma carência de aplicações que promovam a integração e a automatização da operação destas ferramentas, bem como permitam a incorporação de ontologias de domínio ao processo de mapeamento entre a Base de Dados Relacional (BDR) e o formato de dados abertos.

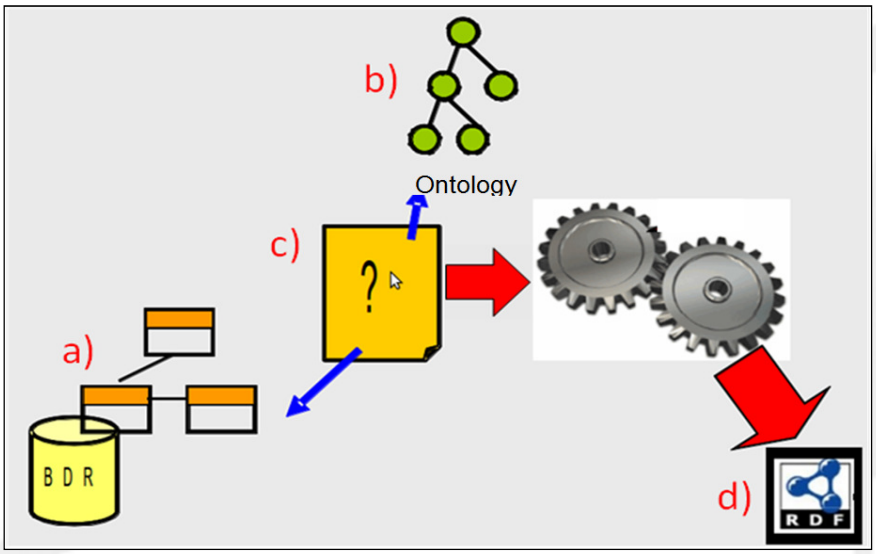
Este artigo tem como objetivo apresentar uma proposta de abordagem para facilitar e automatizar a publicação de dados ligados obtidos a partir de BDRs. Nela são oferecidas uma aplicação com a finalidade de integrar ferramentas de software existentes, e uma nova ferramenta paraa customização semi-automática do arquivo de mapeamento entre a BDR e o formato de dados abertos, de forma a possibilitarque seja nele incorporada uma ontologia de domínio fornecida pelo usuário.

2Referencial Teórico e Trabalhos Relacionados

A *Web* atual, de uma forma geral, tem seu conteúdo destinado à leitura humana (foco no nível de apresentação dos dados), o que impossibilita sua manipulação por computadores. Com a *Web* Semântica, pretende-se que as informações disponibilizadas na *Web* estejam em um formato estruturado, com significados bem definidos, habilitando os computadores a processar e entender automaticamente estes dados (BERNERS-LEE; HENDLER; LASSILA, 2001). Isso promoverá uma mudança da atual "*Web* de documentos" para uma "*Web* de dados" (ANTONIOU; HARMELEN, 2008). Os principais padrões e artefatos atualmente disponíveis para a *Web* Semântica são o modelo de dados *RDF*, as ontologias (lista de termos e relações que descrevem formalmente um domínio de conhecimento), os dados abertos ligados, e a linguagem de consulta *SPARQL*.

O processo de conversão dos dados de uma BDR para um *datasetRDF*, de forma que possa ser publicado na *Web* Semântica no formato de dados abertos ligados, está ilustrado na Figura 1.

Figura 1 – Processo de conversão de uma BDR para um *dataset RDF*



Fonte: Autor

O processo apresentado na Figura 1 pode ser entendido da seguinte forma: a partir da BDR (a), e com base em uma ontologia (b), é gerado (de forma manual ou automática) um arquivo de mapeamento (c), escrito em uma linguagem específica, no qual são relacionados os mapeamentos entre as tabelas da BDR e as classes da ontologia, e entre as colunas da BDR e as propriedades da ontologia; e, com base no arquivo de mapeamento gerado é efetuada, por meio de ferramentas específicas, a transformação de cada linha da BDR em triplas *RDF* (d).

O mapeamento entre BDR e *RDF* pode atualmente ser efetuado por meio de duas técnicas: o mapeamento estático ou o mapeamento dinâmico. O mapeamento estático, também chamado de *RDF dump*, é uma técnica baseada em *ETL* (*Extract, Transform and Load*), onde um repositório *RDF* é criado a partir da conversão de todas as tabelas e colunas de uma BDR, com base em um arquivo de mapeamento. Já o mapeamento dinâmico efetua a conversão das tabelas e colunas de uma BDR para o formato de triplas *RDF* sob demanda (*on-line*), ou seja, de acordo com as respostas a serem retornadas às consultas elaboradas pelo usuário ou pelas ferramentas de publicação de dados abertos ligados (ZHOU, 2010).

A seguir é apresentada uma análise das ferramentas para publicação de dados abertos ligados propostas em alguns trabalhos relacionados e, ao final, é mostrado um comparativo entre estas e a ferramenta desenvolvida na abordagem proposta neste artigo.

A ferramenta *DB2OWL* (CULLOT; GHAWI; Y'ETONGNON, 2007) tem por objetivo converter automaticamente um esquema de BDR em uma ontologia *OWL*, por meio do mapeamento de tabelas em classes e de colunas em propriedades. Já as subclasses e os *object properties* são mapeados a partir dos relacionamentos entre as tabelas. Os mapeamentos gerados são registrados em um arquivo utilizando a linguagem *R2O* (BARRASA; CORCHO; GÓMEZ-PÉREZ, 2004). Algumas limitações da ferramenta são sua compatibilidade restrita aos Sistemas Gerenciadores de Bancos de Dados (SGBDs)*Oracle* e *MySQL*, a impossibilidade de integração de uma ontologia pré-existente ao mapeamento, e a ausência de interface para visualização e consulta dos dados ligados gerados.

O trabalho apresentado por SZEKELY, HEJJA e BUCHMANN (2011) descreve uma aplicação *Web* desenvolvida em linguagem*PHP*, utilizando a *API RAP* (RDF API for PHP), que efetua o mapeamento para *RDF* de uma BDR da área de recursos humanos de uma empresa específica e possibilita a visualização e consulta dinâmica dos dados ligados gerados. O aplicativo utiliza a plataforma *D2RQ*(BIZER; CYGANIAC, 2006) para o mapeamento e a

visualização e consulta de dados ligados. Além de ter escopo muito limitado, o aplicativo não permite a customização do mapeamento ou a incorporação de uma ontologia.

A geração automática de mapeamentos entre BDR e *RDF* utilizando as linguagens *R2RML* ou *Direct Mapping* é o propósito do *plugin* para ambiente *Eclipse* apresentado por SALAS et al. (2011). São oferecidos ao usuário diferentes algoritmos de mapeamento, os quais podem ser escolhidos por meio de uma interface gráfica, bem como outros algoritmos podem ser adicionados. A ferramenta, entretanto, não possibilita a customização do mapeamento ou a incorporação de uma ontologia a este, o que deve ser feito manualmente pelo usuário por meio da edição do arquivo de mapeamento gerado.

O *Triplify*(AUER et al., 2009) é um pequeno *plugin*, desenvolvido em linguagem *PHP*, com a finalidade de converter em tuplas os dados relacionais manipulados por aplicações e páginas da *Web* existentes, e publicá-los na forma de dados ligados (*linked data*). A conversão dos dados relacionais para *RDF* é efetuada mediante a reescrita, na forma de triplas (tabelas como classes e colunas como *URI*), dos resultados provenientes de consultas *SQL* efetuadas nas respectivas BDRs. Dessa forma, uma limitação da ferramenta é que ela não possui suporte a consultas em linguagem *SPARQL*.

O *framework Iconomy* (VAVLIAKIS et al., 2011)tem por objetivo fazer a transformação de BDRs em dados ligados por meio da criação automática de instâncias em uma ontologia *OWL* existente. Oferece um conjunto de interfaces gráficas, por meio das quais é possível: customizar e editar o mapeamento entre uma BDR e uma ontologia em formato *OWL* para criação das instâncias; a construção, de forma interativa, de consultas *SPARQL*; e a inserção de restrições e regras na ontologia, de forma a proporcionar suporte à inferência. As limitações da ferramenta ficam por conta da compatibilidade restrita aos SGBDs *Oracle* e *MySQL*, e do tipo de mapeamento gerado, que é estático (baseado em *ETL*). Além disso, a formulação de consultas *SPARQL* fica limitada às opções oferecidas pela interface gráfica.

O *LOD2 Stack* (AUER et al., 2012) é um *framework* que agrupa uma série de ferramentas para a publicação de dados ligados, abrangendo as tarefas de extração, consulta e exploração, criação, descoberta semi-automática de ligações (*links*) entre fontes de dados e, enriquecimento e reparação de bases de conhecimento. Uma característica relevante do *framework* é a adoção da plataforma *D2RQ*(BIZER; CYGANIAC, 2006) para a extração de dados ligados a partir de BDRs, bem como para a visualização e consultas. Como limitações, não é possível a incorporação de ontologia ao mapeamento entre BDR e *RDF*, e a aplicação está disponível somente para plataforma *Linux*. Além disso, a maior parte das ferramentas oferecidas no *framework* são aplicações disponíveis somente de forma *on-line* na *Web*.

O *Dartgrid*(WU et al., 2006) é um *framework* que oferece um conjunto de ferramentas com o propósito de integrar BDRs heterogêneas usando a *Web* Semântica. Uma das ferramentas oferecidas permite ao usuário, por meio de interface gráfica, construir mapeamentos entre oesquema de uma BDR e uma ontologia. Outra ferramenta possibilita, também por interface gráfica, gerar consultas *SPARQL* baseadas nos termos da ontologia mapeada. Os aspectos negativos destas ferramentas ficam por conta de sua interface gráfica, que está em idioma chinês, e da impossibilidade de edição manual do arquivo de mapeamento gerado. Além disso, o *framework* foi projetado para um domínio específico, que consiste na interconexão de várias BDRs referentes à Medicina Tradicional Chinesa.

Por fim, o *Ne-on Toolkit*(HAASE, et al., 2007) é um *framework* para a construção e o gerenciamento de ontologias que conta com uma grande quantidade de ferramentas e *plugins* para estas finalidades. Apesar de ter seu foco na construção de ontologias, o *framework* conta com um *plugin* específico, chamado de *ODE Mapster*(RODRIGUEZ; GÓMES-PÉREZ, 2006), que oferece uma interface gráfica para o mapeamento semi-automático entre ontologias e BDRs, expresso em linguagem *R2O* (BARRASA; CORCHO; GÓMEZ-PÉREZ,

2004). A aplicação é compatível somente com os SGBDs *Oracle* e *MySQL*, e não oferece interface para a publicação (geração e visualização e consulta) dos dados ligados mapeados.

O diferencial da abordagem proposta neste artigo, em comparação aos trabalhos relacionados, está na automatização, implementada pela interface gráfica, das ferramentas para a geração e customização do mapeamento entre BDR e *RDF*, e para a visualização e consulta dos dados ligados obtidos a partir desse mapeamento. A Tabela 1 apresenta um comparativo entre a abordagem proposta,chamada *RDB2LOD*, e os trabalhos relacionados.

TABELA 1–Comparativo entre a Abordagem RDB2LOD e Trabalhos Relacionados.

QUESITO	DB2OWL	Web App Recursos Humanos	Plugin Eclipse	Triplify	Iconomy	LOD2 Stack	DartGrgrid	Ne-On Toolkit	Abordagem RDB2LOD
Método de Mapeamento	Dinâmico/ Estático	Dinâmico	Dinâmico/ Estático	Dinâmico/ Estático	Estático	Dinâmico/ Estático	Dinâmico/ Estático	Dinâmico/ Estático	**Dinâmico/ Estático**
Permite mapeamentos genéricos?	Sim	Não	Sim	Sim	Sim	Sim	Não	Sim	**Sim**
Linguagem do arquivo de mapeamento	R2O	D2RQ	R2RML/ Direct Mapping	SQL Views	Específica da Aplicação	D2RQ	Específica da Aplicação	R2O	**D2RQ**
SGBDs Compatíveis	MySQL/ Oracle	MySQL/ Oracle	MySQL/ Oracle	Diversos	MySQL/ Oracle	MySQL/ Oracle	MySQL/ Oracle	MySQL/ Oracle	**MySQL/MS-SQL Server**
Ambiente de Execução	Não Informado	Web	Eclipse	Web	Java Runtime	Linux (Debian) / Web	Java Runtime	Java Runtime	**Java Runtime**
Arquivo de mapeamento editável?	Sim	Não	Sim	Sim	Sim	Sim	Não	Sim	**Sim**
Incorporação de ontologia pelo usuário?	Não	Não	Não	Não	Sim	Não	Não	Sim	**Sim**
Visualização e consulta dos dados gerados	Não	Sim	Não	Sim	Sim	Sim	Sim	Não	**Sim**

Fonte: Autor

Destaca-se também como diferenciais dessa abordagem, a compatibilidade com o SGBD *MS SQL Server*, e a ferramenta de implementação do método para incorporação, de forma semi-automática (também por meio de interface gráfica), de uma ontologia ao arquivo de mapeamento, esta compatível com as recomendações de padrão do *W3C*.

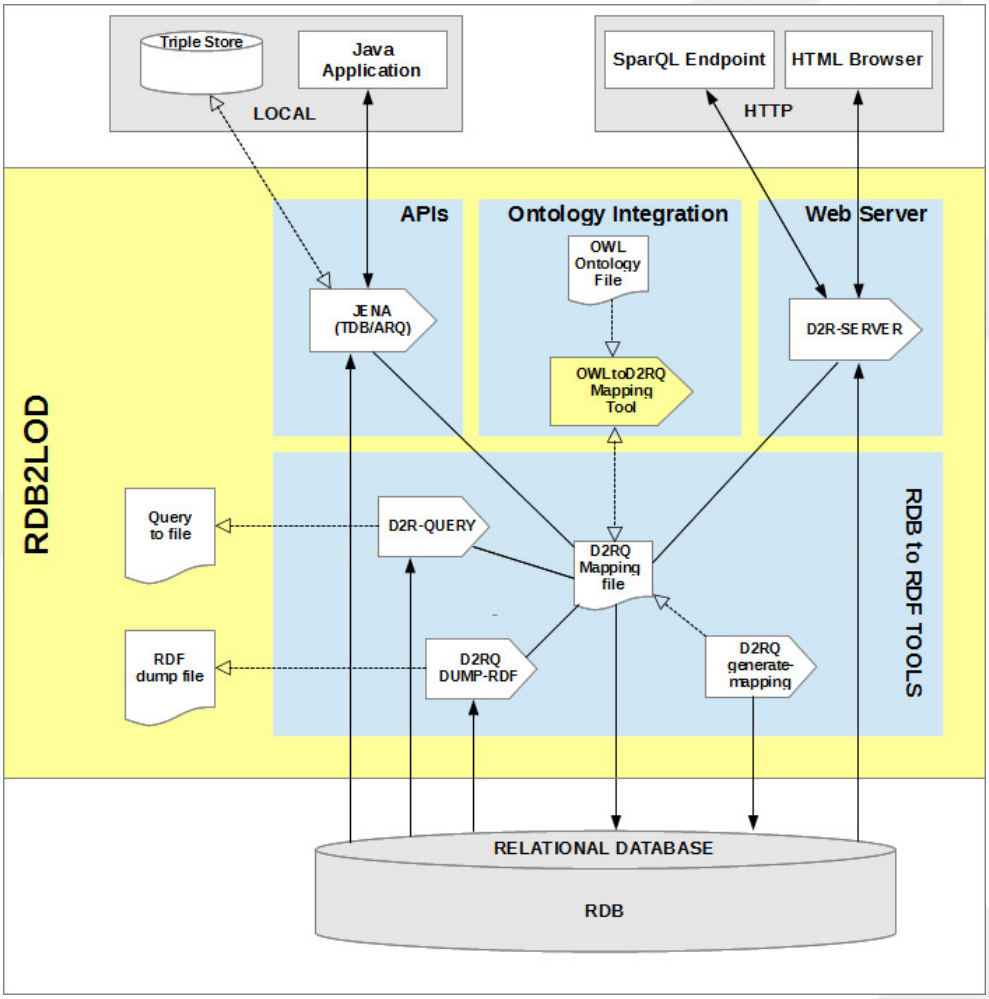
3Abordagem para a publicação de Dados Ligados obtidos a partir de BDRs

As ferramentas atualmente disponíveis para efetuar a conversão de dados mantidos em BDRs para dados ligados na forma de triplas *RDF* possuem diversas limitações, que exigem do usuário um elevado nível de conhecimento técnico e de interação manual. Uma dessas limitações é que várias destas ferramentas são configuradas e acionadas através de linha de comando do sistema operacional. Outra limitação é a necessidade de edição manual do arquivo de mapeamento entre BDR e *RDF* para a substituição do vocabulário padrão (nomes de tabelas como classes e nomes de colunas como propriedades) por elementos de uma ontologia de domínio fornecida pelo usuário.

A abordagem apresentada neste artigo, chamada de *RDB2LOD*, busca suprir essas limitações ao oferecer uma aplicação que integra alguma destas ferramentas, por meio de uma interface gráfica onde é possível configurá-las e acioná-las, e de uma nova ferramenta que possibilita a customização do arquivo de mapeamento para incorporação de uma ontologia do domínio fornecida pelo usuário. Dessa forma, ficam bastante reduzidas as necessidades de

interação manual e de conhecimento técnico por parte do usuário. A Figura 2 apresenta uma visão geral da abordagem *RDB2LOD*, a qual é composta de duas camadas.

Figura 2 – Visão geral da abordagem *RDB2LOD*



Fonte: Autor

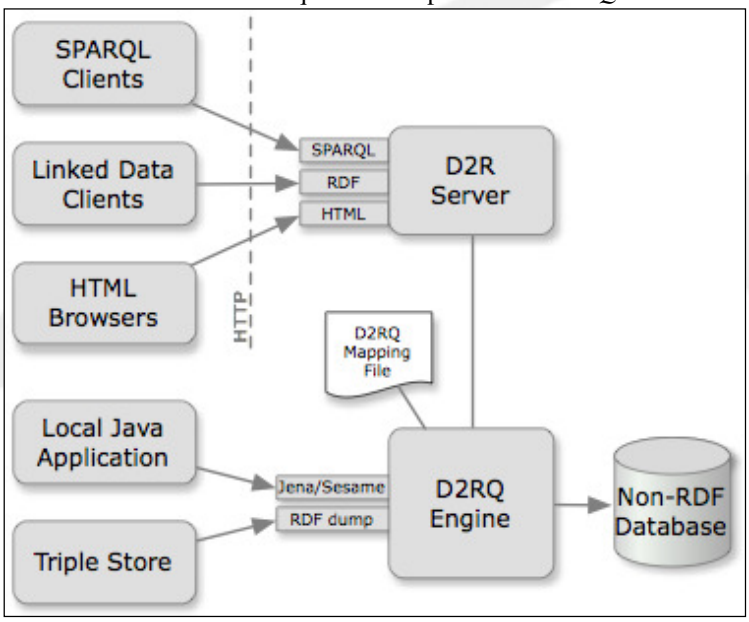
A camada inferior da abordagem conta com uma ferramenta para efetuar a conexão com a BDR, conforme parâmetros fornecidos pelo usuário. Por meio desta ferramenta é obtida a estrutura da BDR e gerado um arquivo de mapeamento, a partir do qual os dados da BDR serão convertidos para o formato de triplas *RDF*.As triplas geradas podem então ser utilizadas por outras duas ferramentas, oferecidas também nesta camada, para a geração de *datasets*, ou então serem repassadas para as ferramentas da camada superior.

Para a seleção de tais ferramentas, dentre os critérios utilizados no processo de avaliação, destacam-se: a possibilidade de edição manual do arquivo de mapeamento gerado (para efetuar a customização e incorporação de ontologia); linguagem de mapeamento empregada pela ferramenta compatível com as linguagens indicadas como recomendação de padrão pelo *W3C* (*R2RML* e *Direct Mapping*); a possibilidade de efetuar mapeamentos tanto dinâmicos (*on-line*) quanto estáticos (baseados em *ETL*); a compatibilidade da ferramenta com os SGBDs *MySQL* e *MS SQL Server*; e, a ferramentaser distribuída na forma de licença pública (*open source/GPL*).

Assim, a plataforma *D2RQ* (BIZER; CYGANIAC, 2006) foi a escolhida, por atender a todos estes critérios. Ela oferece um conjunto de ferramentas para a geração de dados ligados, na forma de triplas *RDF*, a partir do acesso a BDRs. Seus principais componentes são a *D2RQ Mapping Language*, o *D2RQ Engine* e o *D2R Server*. Por meio dessas ferramentas é possível

acessar o conteúdo de BDRs na forma de dados ligados (*Linked Data*), e efetuar consultas *SPARQL* sobre estes dados, localmente ou pela *Web*. Também possibilita a geração de *datasets* em formato *RDF* a partir da descarga de todo o conteúdo (*dump*) da BDR acessada.Uma visão geral de sua arquitetura é apresentada na Figura 3.

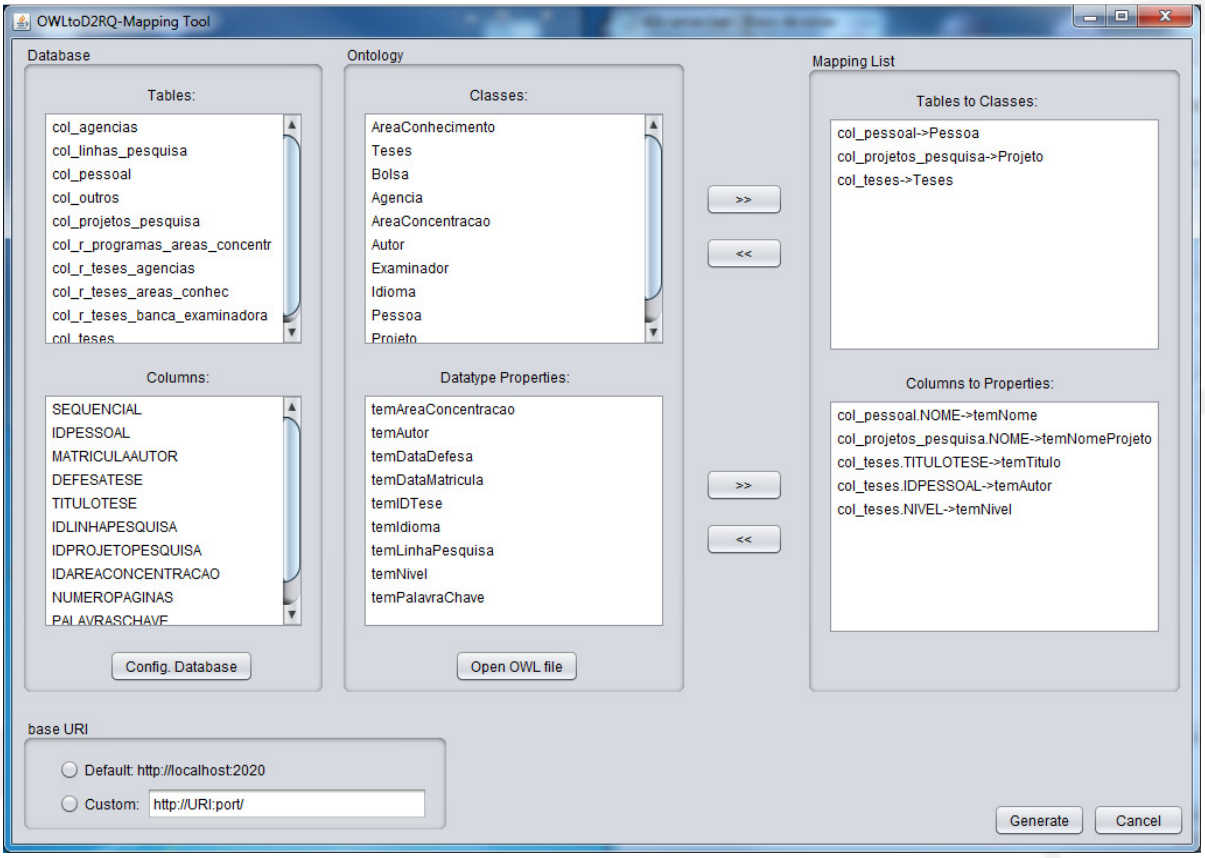
FIGURA 3- Arquitetura da plataforma *D2RQ*.



Fonte: BIZER; CYGANIAC, 2006.

Já a camada superior conta com três componentes distintos: a ferramenta *D2R-Server*, que possibilita a publicação do conteúdo de uma BDR na *Web* Semântica de forma dinâmica, ou seja, efetuando sob demanda a conversão dos dados da BDR mapeada para o formato de triplas *RDF*, sem a necessidade de replicar o conteúdo dessa base de dados em um armazenamento dedicado de dados ligados (*triple store*); as *APIs* do *framework Jena*, com o intuito de possibilitar o desenvolvimento de aplicações externas pelo usuário; e a ferramenta *OWLtoD2RQ-Mapping*, desenvolvida nesta abordagem, que tem por finalidade customizar, de forma interativa e semi-automática, o arquivo de mapeamento entre BDR e *RDF*de forma a permitir a incorporação de uma ontologia do domínio fornecida pelo usuário. A Figura 4 apresenta a interface gráfica desta ferramenta.

Figura 4 – Interface gráfica da ferramenta *OWLtoD2RQ-Mapping* desenvolvida na abordagem



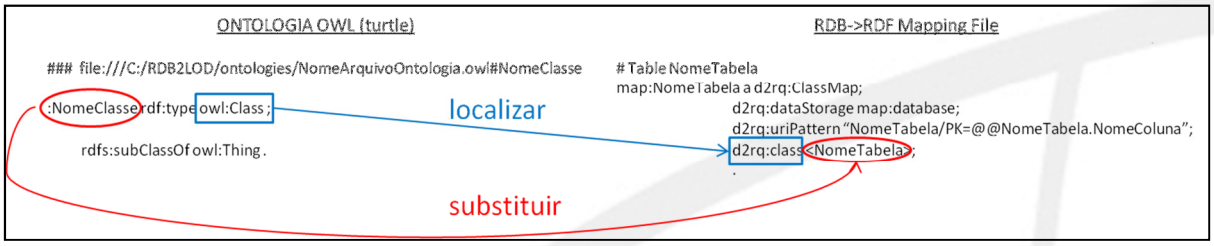
Fonte: Autor

O métodopara efetuar a incorporação dos elementos de uma ontologia *RDFS* ao arquivo de mapeamento entre BDR e *RDF*, implementado pela ferramenta *OWLtoD2RQ-Mapping*, consiste basicamente da edição deste arquivo em três etapas.

Na primeira etapa são inseridas algumas linhas (fixas) no final do cabeçalho do arquivo de mapeamento (cláusulas *@prefix*), para a inclusão de parâmetros relacionados à customização a ser efetuada.

Na segunda etapa são substituídos, no corpo do arquivo de mapeamento, os rótulos das classes mapeadas (que, por padrão, são os nomes das respectivas tabelas do BDR) pelas classes da ontologia *RDFS* associadas, pelo usuário, por meio da interface gráfica da ferramenta. Para cada substituição a ser processada no arquivo de mapeamento, é efetuada a localização da cláusula "*d2rq:class*", acompanhada do nome da respectiva tabela, e, quando encontrada, esta é substituída pelo nome da classe da ontologia que foi associada a esta tabela.A Figura 5 permite uma representação deste procedimento.

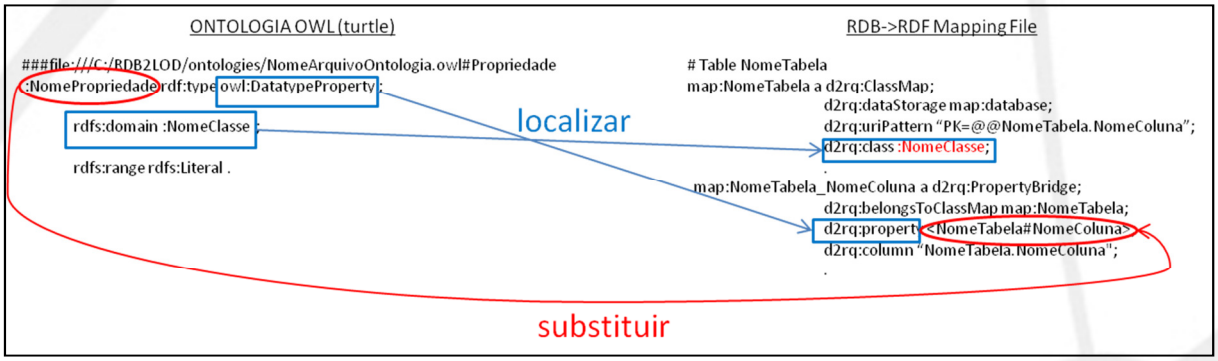
Figura 5 – Localização e substituição do nome da tabela pela classe da ontologia *RDFS* associada



Fonte: Autor

Na terceira etapa, são localizados e substituídos, para cada classe (tabela BDR) no arquivo de mapeamento, os rótulos das propriedades, ou seja, os nomes das colunas BDR mapeadas. Para cada propriedade da ontologia *RDFS* fornecida, são localizados no arquivo de mapeamento a correspondente classe (cláusula *d2rq:class*) e a respectiva coluna da tabela BDR associada a esta propriedade (cláusula *d2rq:property*) para, então, efetuar a substituição do nome da coluna BDR pelo nome da propriedade *RDFS* associada. A Figura 6 permite visualizar esta alteração.

Figura 6 – Localização e substituição do nome da coluna BDR pela propriedade *RDFS* associada

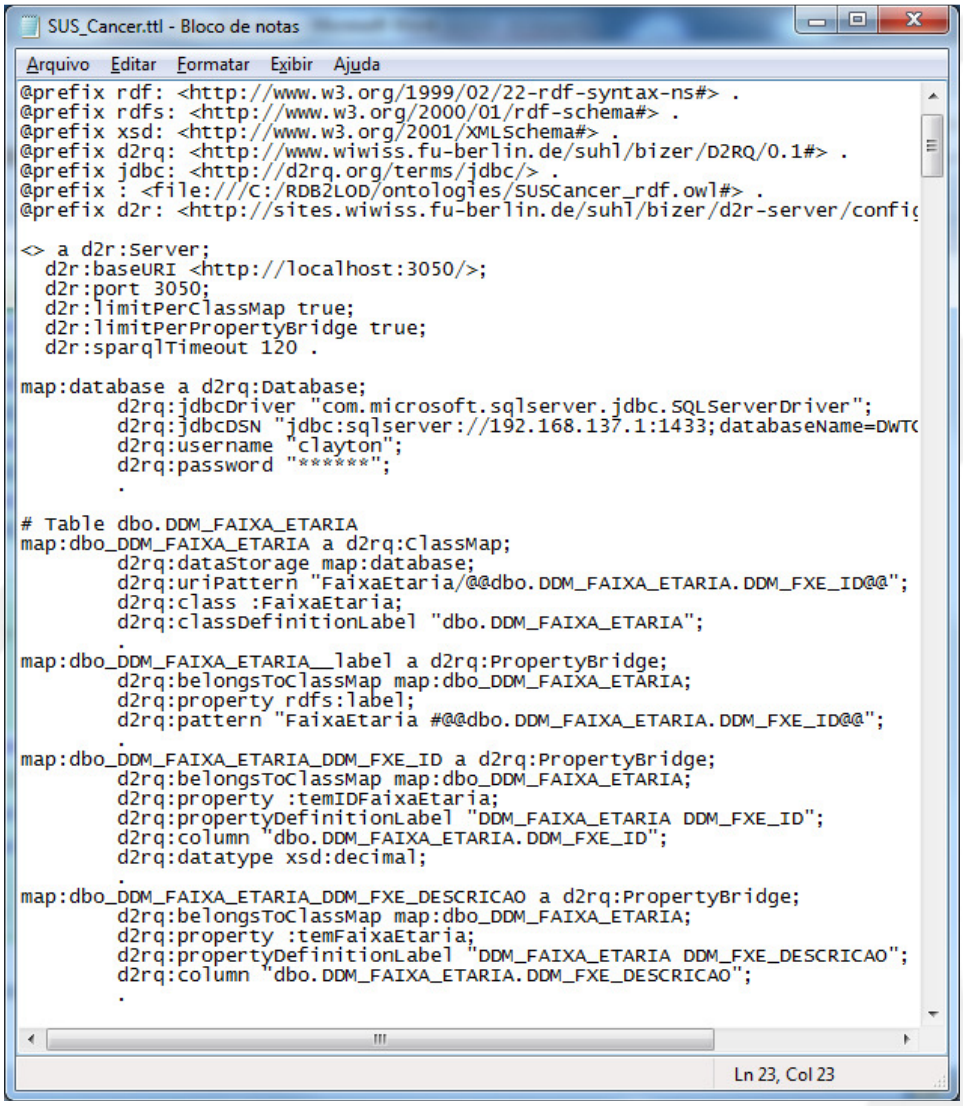


Fonte: Autor

Ao final dessas três etapas de alteração, o arquivo de mapeamento entre BDR e *RDF*, já com os elementos da ontologia *RDFS* incorporados, pode ser utilizado para gerar os dados ligados na forma de triplas *RDF*. Assim, o valor (ou objeto) de cada tripla *RDF* a ser gerada corresponderá a um campo de uma coluna (predicado) referente a uma tabela (sujeito) da BDR mapeada, ou ainda ao campo chave de outra tabela relacionada (expressa por meio de uma *URI*), quando esta coluna (predicado) for uma chave estrangeira da tabela mapeada. Este arquivo customizado de mapeamento pode ser visualizado na Figura 7.

A partir desse método será possível proporcionar um ganho de expressividade às triplas *RDF* geradas, ao apresentar significados bem definidos para os sujeitos, predicados e objetos destas, obtidos da customização do mapeamento da BDR com a incorporação da ontologia do respectivo domínio.

Figura 7 – Arquivo customizado de mapeamento de uma BDR para o modelo de dados *RDF*.

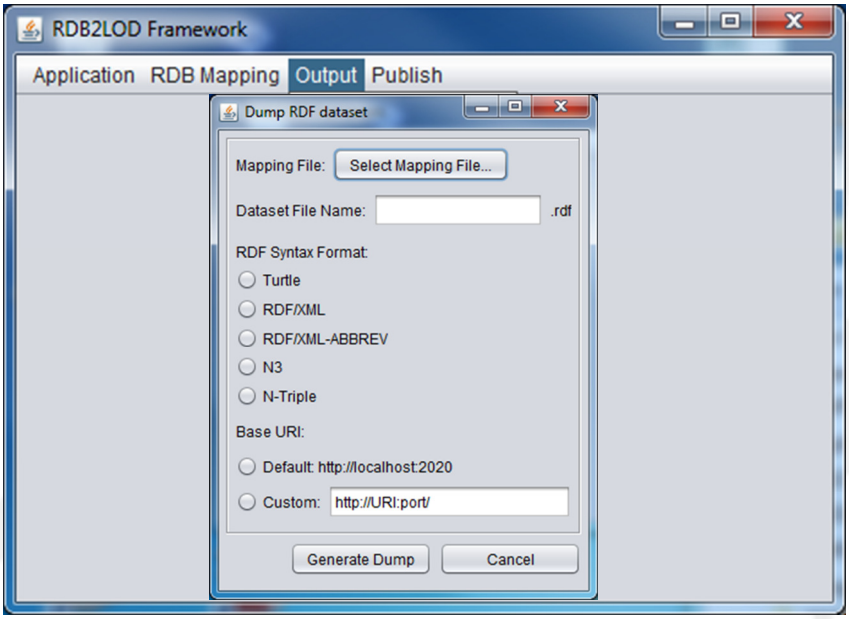


Fonte: Autor

Para prover as interfaces gráficas necessárias para a automatização e integração das ferramentas que compõem a abordagem, foi desenvolvido um aplicativo, em linguagem *Java*, chamado *RDB2LOD*. Este aplicativo possibilita a coleta de parâmetros, por meio de campos e listas ou botões de seleção, com a finalidade de automatizar a elaboração e execução das linhas de comando para acionamento e operação de cada uma das ferramentas aplicadas na abordagem.

Para cada uma das opções do menu principal do aplicativo, é aberta uma janela para a coleta de parâmetros que são armazenados em variáveis locais do código *Java*, de forma a permitir a construção e execução automatizadas da linha de comando da respectiva ferramenta. A Figura 8 mostra uma tela do aplicativo.

Figura 8 – Tela do aplicativo desenvolvido para integração das ferramentas aplicadas na abordagem.



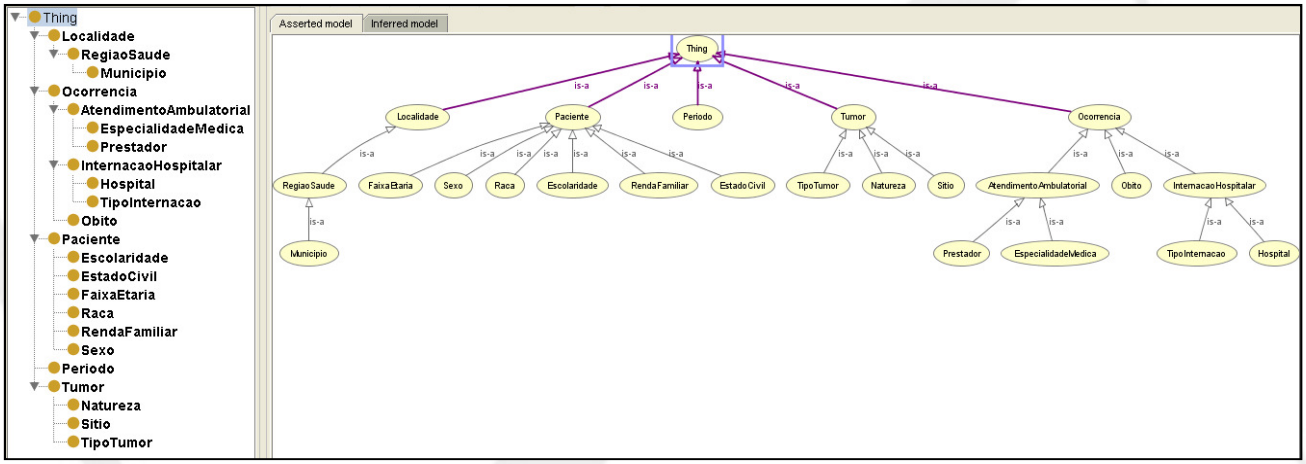
Fonte: Autor

4Estudo de Caso

A abordagem *RDB2LOD* foi aplicada a um estudo de caso em que foi utilizada uma BDR que contém dados referentes ao acompanhamento (mortes, internações hospitalares e atendimentos ambulatoriais) de portadores de câncer, a qual é mantida pelo Sistema Único de Saúde – SUS, e encontra-se na forma de um *datawarehouse* (chamado DW-SUS), onde são armazenados os dados históricos coletados de três sistemas de informações distintos: SIM (Sistema de Informações sobre Mortalidade); SIA (Sistema de Informações sobre Atendimento Ambulatorial); e SIH (Sistema de Informações Hospitalares).

Para a aplicação da abordagem, com o propósito de gerar e publicar os dados ligados obtidos a partir do mapeamento customizado dessa BDR, primeiramente foi obtida uma ontologia do domínio "Acompanhamento de Tumores Cerebrais", mostrada na Figura 9.

Figura 9 – Tela do aplicativo desenvolvido para integração das ferramentas aplicadas na abordagem

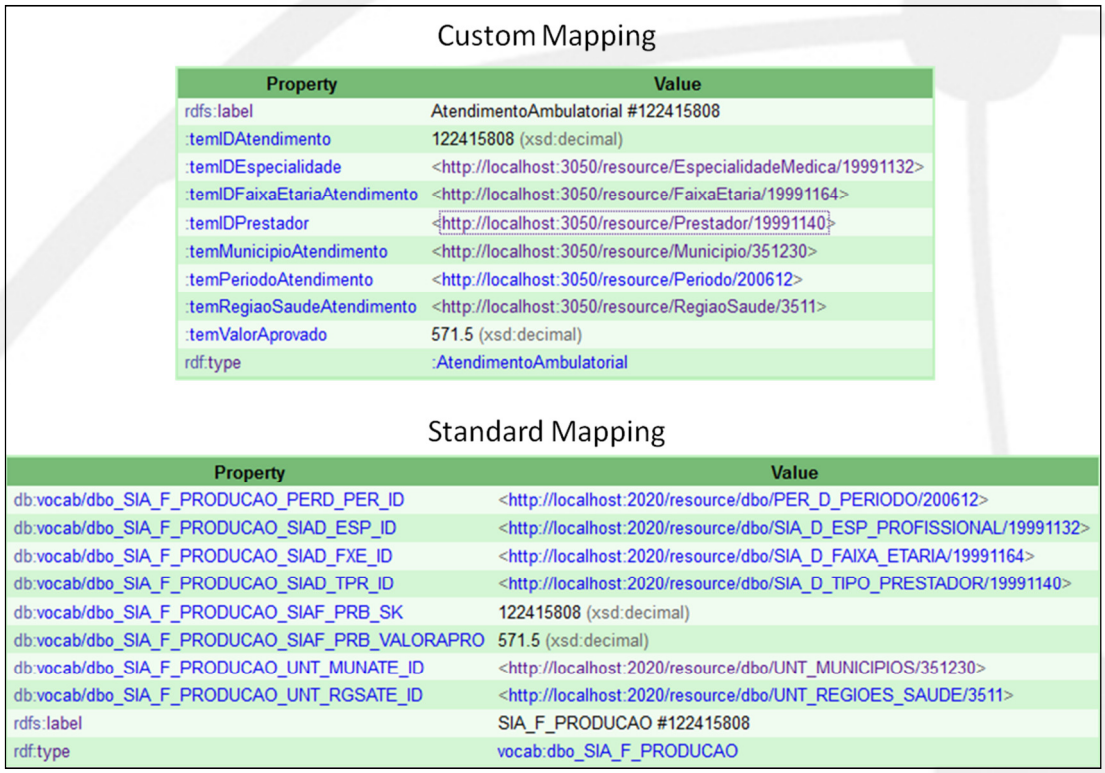


Fonte: Autor

A partir do arquivo dessa ontologia *RDFS* (em formato *OWL*), e dos parâmetros para conexão à BDR, foi possível efetuar a customização do arquivo de mapeamento, de forma semi-automática, por meio da ferramenta *OWLtoD2RQ-Mapping*, que efetuou a substituição do vocabulário padrão pelos termos da ontologia do domínio.

Em seguida, foram comparadas as visualizações dos dados ligados gerados, tanto a partir do mapeamento padrão da BDR, quanto a partir do mapeamento customizado (com a incorporação da ontologia do domínio). A Figura 10 apresenta uma visualização de dados referentes a uma internação hospitalar.

Figura 10 – Comparação entre visualizações de dados referentes a uma internação hospitalar, efetuadas com e sem a customização do arquivo de mapeamento.



Custom Mapping

Property	Value
rdfs:label	AtendimentoAmbulatorial #122415808
:temIDAtendimento	122415808 (xsd:decimal)
:temIDEspecialidade	<http://localhost:3050/resource/EspecialidadeMedica/19991132>
:temIDFaixaEtariaAtendimento	<http://localhost:3050/resource/FaixaEtaria/19991164>
:temIDPrestador	<http://localhost:3050/resource/Prestador/19991140>
:temMunicipioAtendimento	<http://localhost:3050/resource/Municipio/351230>
:temPeriodoAtendimento	<http://localhost:3050/resource/Periodo/200612>
:temRegiaoSaudeAtendimento	<http://localhost:3050/resource/RegiaoSaude/3511>
:temValorAprovado	571.5 (xsd:decimal)
rdf:type	:AtendimentoAmbulatorial

Standard Mapping

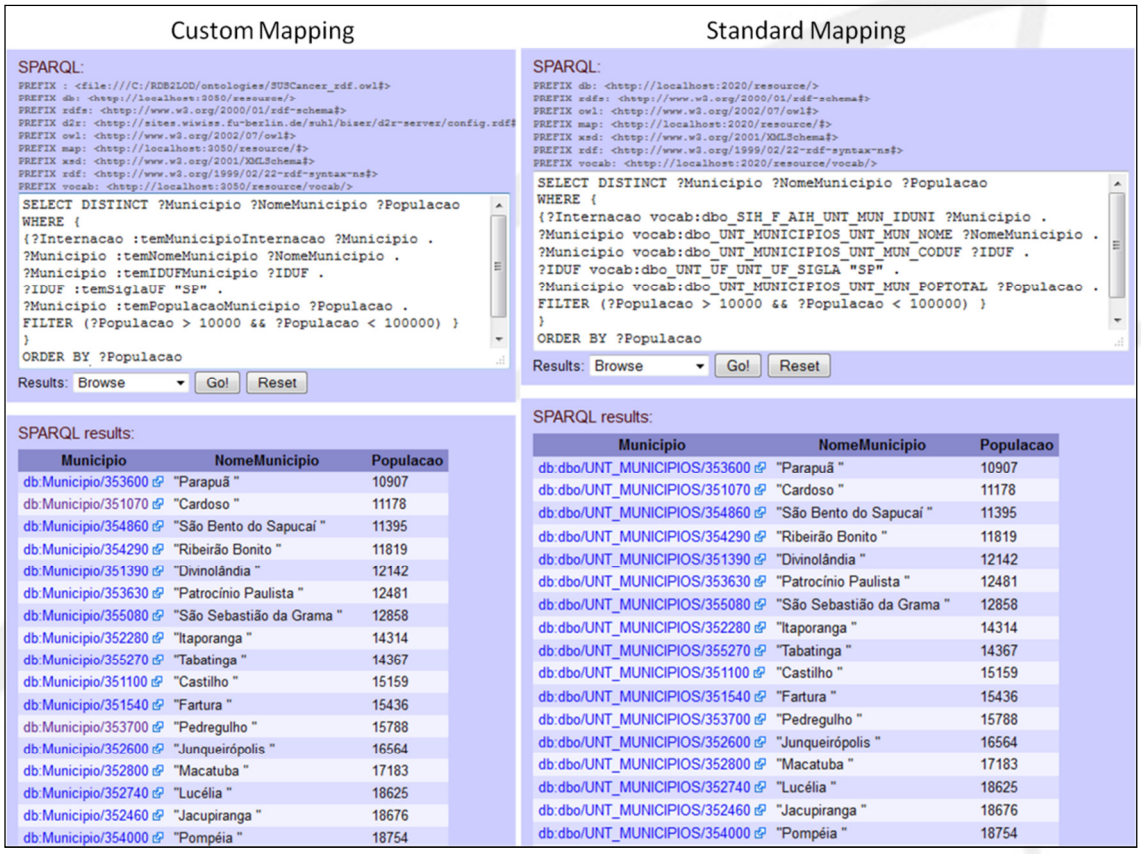
Property	Value
db:vocab/dbo_SIA_F_PRODUCAO_PERD_PER_ID	<http://localhost:2020/resource/dbo/PER_D_PERIODO/200612>
db:vocab/dbo_SIA_F_PRODUCAO_SIAD_ESP_ID	<http://localhost:2020/resource/dbo/SIA_D_ESP_PROFISSIONAL/19991132>
db:vocab/dbo_SIA_F_PRODUCAO_SIAD_FXE_ID	<http://localhost:2020/resource/dbo/SIA_D_FAIXA_ETARIA/19991164>
db:vocab/dbo_SIA_F_PRODUCAO_SIAD_TPR_ID	<http://localhost:2020/resource/dbo/SIA_D_TIPO_PRESTADOR/19991140>
db:vocab/dbo_SIA_F_PRODUCAO_SIAF_PRB_SK	122415808 (xsd:decimal)
db:vocab/dbo_SIA_F_PRODUCAO_SIAF_PRB_VALORAPRO	571.5 (xsd:decimal)
db:vocab/dbo_SIA_F_PRODUCAO_UNT_MUNATE_ID	<http://localhost:2020/resource/dbo/UNT_MUNICIPIOS/351230>
db:vocab/dbo_SIA_F_PRODUCAO_UNT_RGSATE_ID	<http://localhost:2020/resource/dbo/UNT_REGIOES_SAUDE/3511>
rdfs:label	SIA_F_PRODUCAO #122415808
rdf:type	vocab:dbo_SIA_F_PRODUCAO

Fonte: Autor

Na Figura 10 é possível perceber que os rótulos das propriedades, quando gerados a partir do mapeamento customizado da BDR, apresentam um significado bem definido (aproximados à linguagem natural) para os respectivos valores apresentados, ao contrário dos rótulos de propriedades gerados a partir do mapeamento padrão, onde são utilizados nomes das colunas da BDR, que neste caso não impedem, mas dificultam a interpretação dos respectivos valores apresentados, pois dependem da compreensão de seus significados.

Por fim, foi efetuada uma comparação entre uma consulta *SPARQL* formulada a partir do mapeamento padrão, e a mesma consulta formulada a partir do mapeamento customizado da BDR. A Figura 11 mostra as consoles *SPARQL* com as consultasformuladas e os respectivos resultados obtidos, para ambos os mapeamentos. É importante observar as diferenças entre as sintaxes das consultas *SPARQL* formuladas, tanto para o mapeamento customizado, quanto para o mapeamento padrão. A utilização dos termos da ontologia do domínio no mapeamento customizado, em substituição aos nomes de colunas BDR no mapeamento padrão, permite uma melhor compreensão das propriedades (predicados das triplas *RDF*), que se aproximam da linguagem natural e facilitam uma eventual correção ou posterior edição dessa consulta.

Figura 11 – Comparação entre consultas *SPARQL*, efetuadas com e sem a customização do arquivo de mapeamento.



Fonte: Autor

A partir das comparações apresentadas, tanto na visualização quanto na consulta *SPARQL*, é possível afirmar que a aplicação da abordagem *RDB2LOD* proporcionou uma melhor expressividade às triplas *RDF* geradas, ao apresentar significados bem definidos (aproximados à linguagem natural) para os sujeitos, predicados e objetos destas, que foram obtidos da customização do mapeamento da *BDR* com a incorporação de uma ontologia do domínio.

A aplicação da abordagem *RDB2LOD* também proporcionou uma interação do usuário por meio de interfaces gráficas, eliminando a necessidade de interação manual para a configuração e operação das ferramentas aplicadas ao processo de publicação de dados ligados. Desta forma, a customização do mapeamento pelas associações entre tabelas e colunas da BDR e as classes e propriedades da ontologia, que até então era feita manualmente, passa a ser feita de forma automatizada por meio desta abordagem, o que reduz o tempo gasto e a necessidade de conhecimento técnico da linguagem do arquivo de mapeamento gerado pela ferramenta aplicada. A customização do mapeamento de bases de dados com um número elevado de tabelas e colunas, que seria impraticável pelo método convencional (manualmente), passa a ser possível por meio desta abordagem.

5Conclusão

Estaabordagem vem preencher uma lacuna encontrada nas ferramentas disponíveis para a geração e publicação de dados ligados a partir de dados estruturados persistidos em bases de dados relacionais, que é a falta de uma interface gráfica que automatize todo o processo e integre as diferentes ferramentas aplicadas, bem como gerar as triplas *RDF* com

base em uma ontologia de domínio, de forma a conferir maior expressividade às visualizações e consultas desses dados.

A abordagem *RDB2LOD* possibilita a integração de ferramentas e a automatização do processo de publicação de dados ligados obtidos a partir de BDRs, ao oferecer uma interface gráfica que reduz a necessidade de conhecimento técnico e elimina a necessidade de interação manual do usuário.

A partir do método e da ferramenta gráfica para customização do mapeamento entre a BDR e o formato *RDF*, com a incorporação de uma ontologia do domínio foi possível proporcionar melhor expressividade às triplas *RDF* geradas, que passaram a apresentar significados bem definidos (aproximados à linguagem natural) para os sujeitos, predicados e objetos. Isso impactou de forma positiva a visualização dos dados ligados publicados, a geração de *datasets*, e a elaboração de consultas *SPARQL*. Da mesma forma, a aplicação de uma ontologia do domínio nesta customização pode evitar a geração de triplas irrelevantes ou desnecessárias, possibilitando melhor visualização e exploração dos dados ligados obtidos.

Como trabalhos futuros, que possam complementar ou expandir a abordagem RDB2LOD, são sugeridos: dotar a ferramenta *OWLtoD2RQ-Mapping* de capacidade para gerar e customizar o mapeamento a partir de múltiplas BDRs, bem como ampliar a compatibilidade com outros SGBDs; permitir a associação entre tabelas da BDR e *Object Properties* da ontologia em formato *OWL* a ser incorporada na customização do mapeamento, de forma a estabelecer outras relações que não estavam implícitas no esquema desta BDR; e, possibilitar o mapeamento de regras, restrições e cardinalidade para as classes, durante a customização do mapeamento da BDR.

Para viabilizar a distribuição do aplicativo *RDB2LOD*na forma de licença pública (*open source/GPL*), recomenda-se ainda como trabalho futuro a correção e otimização de seu código-fonte (escrito em linguagem *Java*). O protótipo de aplicativo desenvolvido para validação da abordagem foi customizado para execução em ambiente *MS-Windows*. A execução em ambiente *Linux* pode requerer nova customização do aplicativo.

Por fim conclui-se que a abordagem *RDB2LOD* permitirá a publicação na *Web*, em grande escala, de dados ligados obtidos a partir de bases de dados relacionais, de forma a possibilitar amplo acesso a muitos usuários, o que hoje ocorre ainda de forma muito tímida por meio de algumas iniciativas.

Referências

ALLEMANG, D.; HENDLER, J. A. **Semantic web for the working ontologist:** modeling in RDF, RDFS and OWL. Burlington: Morgan Kaufmann, 2008.

ANTONIOU, G.; HARMELEN, F. **A semantic web primer.** 2ed. Cambridge: MIT Press, 2008.

AUER, S. et al. Triplify- light-weight linked data publication from relational databases.In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 18., 2009, Madrid. **Proceedings…** New York: ACM, 2009. p. 621-630.Disponível em: <http://doi.acm.org/10.1145/1526709.1526793>. Acesso em: 20 set. 2012.

AUER, S. et al. Managing the life-cycle of linked data with the LOD2 stack.In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 11., 2012, Boston. **Proceedings...** Berlin: Springer, 2012. p. 1-16. Disponível em: <http://lod2.eu/ BlogPost/1214-paper-about-lod2-stack-accepted-for-iswc.html>. Acesso em: 20 set. 2012.

BARRASA, J.; CORCHO, Ó.; GÓMEZ-PÉREZ, A. R2O, an extensible and semantically based database-to-ontology mapping language.In: WORKSHOP ON SEMANTIC WEB AND DATABASES, 2., 2004, Toronto. **Proceedings...**New York: Springer, 2004. p. 1069-1070.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, New York, v. 284, p. 28-37, Mai. 2001.

BIZER, C.; CYGANIAC, R. **D2RQ platform:** publishing relational databases on the semantic web. 2006.Poster presented at the 5th. Conference International Semantic Web, 2006, Athens. Disponível em:<http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/Bizer-Cyganiak-D2R-Server-ISWC2006.pdf>. Acesso em: 01 out. 2012.

CHOI, M.-Y. et al. Interoperability between a relational data model and an RDF data model.In:INTERNATIONAL CONFERENCE ON NETWORKED COMPUTING AND ADVANCED INFORMATION MANAGEMENT, 6., 2010, Seoul. **Proceedings...** Piscataway: IEEE, 2010. p. 335-340.

CULLOT, N.; GHAWI, R.; Y'ETONGNON, K. DB2OWL: a tool for automatic database-to-ontology mapping. In: ITALIAN SYMPOSIUM ON ADVANCED DATABASE SYSTEMS, 15., 2007, Torre Canne. **Proceedings…**Torre Canne: Dipartimento di Informatica, Università degli Studi Di Bari, 2007. p. 491-494. Disponível em: <http://dblp.unitrier.de/rec/bibtex/conf/sebd/CullotGY07>. Acesso em: 20 set. 2012.

HAASE, P. et al. **Ontology engineering and plugin development with the neon toolkit.**2007.Tutorial presented at the 6th International Semantic Web Conference, 2007, Busan.

HERT, M. et al. UpLink: a linked data editor for BDR-to-RDF data.In: INTERNATIONAL CONFERENCE ON SEMANTIC SYSTEMS, 7., 2011, New York. **Proceedings...** New York: ACM, 2011. p. 159-162. Disponível em:<http://doi.acm.org/10.1145/2063518.2063539>. Acesso em: 20 set. 2012.

LING, H.; ZHOU, S. Translating relational databases into RDF.In:INTERNATIONAL CONFERENCE ON ENVIRONMENTAL SCIENCE AND INFORMATION APPLICATION TECHNOLOGY, 2., 2010, Wuhan. **Proceedings...** Piscataway: IEEE, 2010. p. 464-467.

LV, Y.; MA, Z. M. Transformation of relational model to RDF model.In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS, 2008, Singapore. **Proceedings...** Piscataway: IEEE, 2008. p. 506-511.

MACHADO, A. L.; PARENTE DE OLIVEIRA, J. M. DIGO: an open data architecture for e-government.In: ENTERPRISE DISTRIBUTED OBJECT COMPUTING CONFERENCE WORKSHOPS, 15., 2011, Helsinki. **Proceedings…** Piscataway: IEEE, 2011. p. 448-456.

MOHAMED, H.; JINCAI, Y.; QIAN, J. Towards integration rules of mapping from relational databases to semantic web ontology.In:INTERNATIONAL CONFERENCE ON WEB INFORMATION SYSTEMS AND MINING, 2010, Sanya. **Proceedings...** Piscataway: IEEE, 2010. p. 335-339.

RODRIGUEZ, J. B.; GÓMEZ-PÉREZ, A. Upgrading relational legacy data to the semantic web.INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 15., 2006, Edinburgh. **Proceedings...**New York: ACM, 2006. p. 1069-1070. Disponível em: <http://doi.acm.org/10.1145/1135777.1136019>. Acesso em: 20 set. 2012.

SAHOO, S. S. et al. **A survey of current approaches for mapping of relational databases to RDF.** [S.l.]: W3C, 2009. Disponível em:<http://www.w3.org/2005/Incubator/BDR2rdf/BDR2RDF_SurveyReport.pdf>. Acesso em: 20 set. 2012.

SALAS, P. E. et al. BDR2RDF plugin: relational databases to RDF plugin for eclipse.In:WORKSHOP ON DEVELOPING TOOLS AS PLUG-INS, 1., 2011, Honolulu. **Proceedings...**New York: ACM, 2011. p. 28-31. Disponível em: <http://doi.acm.org/10.1145/1984708.1984717>. Acesso em: 20 set. 2012.

SEQUEDA, J. F.; ARENAS, M.; MIRANKER, D. P. On directly mapping relational databases to RDF and OWL.In:INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 21., 2012, Lyon. **Proceedings...** New York: ACM, 2012. Disponível em:<http://doi.acm.org/10.1145/2187836.2187924>. Acesso em: 20 set. 2012.

SZEKELY, A.; HEJJA, A.; BUCHMANN, R. A. Mapping a relational database into a RDF repository.In:INTERNATIONAL SYMPOSIUM ON SYMBOLIC AND NUMERIC ALGORITHMS FOR SCIENTIFIC COMPUTING, 13., 2011, Timisoara. **Proceedings...**Piscataway: IEEE, 2011. p.175-182.

TURBAN, E. et al. **Administração de tecnologia da informação.** Rio de Janeiro: Campus, 2005.

VAVLIAKIS, K. N. et al. An integrated framework for enhancing the semantic transformation, editing and consultaing of relational databases.**Expert Systems with Applications**, Amsterdam, v. 38, n. 4,. p. 3844-3856,apr. 2011.

WU, Z. et al. Dartgrid: a semantic web toolkit for integrating heterogeneous relational databases.In: SEMANTIC WEB CHALLENGE AT 4TH INTERNATIONAL SEMANTIC WEB CONFERENCE, 4th., 2006, Athens. **Proceedings...** [S.l.]: Semantic Web Science Association, 2006. p. 750-763. Disponível em: <http://www.aifb.uni-karlsruhe.de/WBS/ywa/publications/wu06TCM_ISWC06.pdf>. Acesso em: 6 out. 2012.

ZHOU, S. Exposing relational database as RDF.In: INTERNATIONAL CONFERENCE ON INDUSTRIAL AND INFORMATION SYSTEMS, 2., 2010, Dalian. **Proceedings...** Piscataway: IEEE, 2010. p. 237-240.