A.to.Z NOVAS PRÁTICAS EM INFORMAÇÃO E CONHECIMENTO

ARTIGO

Linked Open Data como forma de agregar valor às informações clínicas

Fernando Hadad Zaidan, Marcello Peixoto Bax

Resumo

Introdução: Os dados abertos vinculados (*Linked Open Data* – LOD) têm sido assunto constante nos principais congressos e *journals* de *web* semântica por todo o mundo. Diversos estudos comprovam que o consumo destes dados é potencialmente importante para melhorar a qualidade dos sistemas de informações nas mais diversas áreas. A gestão da informação clínica é uma destas áreas, e a iniciativa LOD tem se esforçado para padronizar a sua publicação, tornando a interligação do conjunto de dados (*datasets*) mais eficiente. Contudo, agregar valor aos dados internos dos sistemas de informação, utilizando o LOD é desafiador. O objetivo deste trabalho é apresentar o LOD, a fim de constatar a possibilidade se agregar valor aos dados de sistemas de informação clínicos. **Método**: Com base nos pesquisadores que propuseram a *web* semântica e os que acompanham a evolução do LOD efetivou-se uma revisão de literatura dos principais conceitos. Uma pesquisa documental foi realizada obtendo uma bibliografia baseada em livros e *papers* de autores renomados. **Resultados**: Foram apresentadas seis tecnologias que utilizam o padrão de dados abertos vinculados e que propõem a agregação de valor aos dados de sistemas de informação, em especial os clínicos. Os pontos em comum destas seis tecnologias situam-se na publicação, extração, interligação e consumo dos dados do LOD. **Conclusões**: Conclui-se que é possível agregar valor aos dados internos dos sistemas clínicos, mesmo quando se tem disponível mais de três bilhões de triplas do LOD para serem utilizadas/consumidas.

Palavras-chave

Web semântica. Sistemas de informação. Dados abertos vinculados.

Introdução

Os sistemas de informação organizacionais têm evoluído e tendem a sofisticação, assumindo um grau de inteligência elevado, e se mostrando adequados para a interligação de dados organizacionais (LAUDON; LAUDON, 2011). Porém, isto não é tudo o que se necessita para agregar valor aos dados. A obtenção de dados ainda é realizada de maneira precária, desconectada e às vezes manual, comprometendo resultados e a integração. Os bancos de dados relacionais, muito eficientes em diversos contextos, não conseguem fornecer a capacidade, por si só, de uma operabilidade integrada em uma *web* distribuída (HARTIG; LANGEGGER, 2010).

A publicação e o consumo de dados na *web*, promovendo a flexibilidade entre os sistemas interconectados, tende a ser facilitado com as tecnologias semânticas. O termo tecnologia semântica é representado por um conjunto de famílias que buscam o significado dos dados e da informação (ALLEMANG; HENDLER, 2011). Em uma *web* mais sofisticada – que privilegie a consistência, significado e a integridade dos dados – torna-se necessário o apoio no nível dos dados. Logo, não se tem uma página apontando para outra, mas um dado apontando para outro dado, usando referências globais (*Uniform Resource Identifiers* – URI[1]).

O termo *Linked Data* foi cunhado por Tim Berners-Lee em 2006 e refere-se a uma *web* que possa conectar dados relacionados, tornando-os mais úteis, e que contribua para diminuição das barreiras para ligar dados (LINKED DATA, 2013). Já o *Linked Open Data* (LOD), traduzido

[1] os conceitos de *Uniform Resource Identifiers* (URI), assim como o de *Resource Description Framework* (RDF), serão fundamentados no item *Web* semântica e seus conceitos.

livremente por dados abertos vinculados, é um projeto aberto comunitário mundial iniciado em 2007 e que visa à publicação de vários conjuntos de dados (*datasets*) de forma que as ligações sejam possíveis entre eles. A responsabilidade deste projeto fica a cargo do *World Wide Web Consortium* (W3C).

Em um novo cenário de interoperabilidade e de dados abertos vinculados torna-se possível um modelo de dados onde a informação sobre uma única entidade é distribuída na *web* e é acessada por inúmeras organizações, o que agrega valor aos sistemas de informação (HEATH; BIZER, 2011, LINKED DATA, 2013). Este modelo de dados, o LOD, fará parte da infraestrutura básica da *web* disponibilizando de forma acessível os dados interligados e acomodando mais facilmente as aplicações que ainda mantém os dados internamente (ALLEMANG; HENDLER, 2011).

Em 2001, quando foram apresentadas as primeiras proposições da *web* semântica por Tim Berners-Lee (BERNERS-LEE; HENDLER; LASSILA, 2001), a *web* era constituída somente de textos e documentos e não possuía dados marcados semanticamente (ALLEMANG; HENDLER, 2011; WORLD WIDE..., 2013). Hoje, na *web* de dados, são encontrados vários domínios de natureza diversa, com um significativo conjunto de dados (*datasets*) no LOD, dentre eles os dados da Wikipedia (cujo *dataset* é o DBPedia), assim como os dados de governos, geográficos, censo, saúde, entretenimento, acadêmicos, cujos *datasets* são, respectivamente, o *Data.Gov.UK*; o *GeoNames*, o *United States* (US) *Census*, o *DailyMed*, o *British Broadcasting Corporation Music* (BBC *Music*), e o *Association for Computing Machinery* (ACM). Não são poucas as iniciativas para um efetivo crescimento de *datasets* do LOD no formato *Resource Description Framework* (RDF) – o modelo de dados da *web* semântica (LINKED DATA, 2013).

Justifica-se este estudo com base na viabilidade de se utilizar os dados disponíveis nestes *datasets* em diversos domínios (LINKED DATA, 2013) e, diante do que foi exposto, formulou-se a seguinte questão de investigação específica para o domínio da saúde: de que forma a utilização dos dados abertos do LOD podem agregar valor aos dados internos dos Sistemas de Informação Clínicos (*Clinic Information Systems* – CIS)?

Assim, o objetivo do presente artigo é apresentar o *Linked Open Data* (LOD), sua evolução e nuvem de dados abertos vinculados a fim de constatar se é possível agregar valor aos dados e informações clínicos. A nuvem de dados é representada por uma imagem contendo os *datasets* com suas interligações, que são publicados no LOD. Efetivou-se uma breve análise sobre a publicação e consumo destes dados abertos, e de que forma eles poderão ser interligados aos dados fechados dos Sistemas de Informação Clínicos (CIS).

Este artigo está dividido em cinco partes. Acima, contextualizou-se o tema e foram apresentados os objetivos, as justificativas e a questão de investigação. Na seção seguinte, os conceitos são elucidados e fundamentados. Na terceira parte descreve-se a metodologia utilizada no intuito de alcançar o objetivo do estudo; a quarta parte do artigo esclarece a forma de agregar valor aos dados internos dos CIS e a apresentação de algumas tecnologias, seguida das considerações finais.

Procedimentos metodológicos

Esta pesquisa tem o carácter exploratório, sobretudo pelo fato de contribuir para tornar o tema central, o LOD, mais explícito à comunidade acadêmica, assim como para possibilitar a evolução dos estudos.

A técnica utilizada foi a pesquisa documental, apoiada na leitura de livros cujos autores são seminais deste tema. Selecionaram-se os últimos trabalhos nos principais congressos e *journals* internacionais da área, cuja pesquisa foi realizada no Portal CAPES nos dois primeiros meses de 2013. Desta forma, obteve uma lista de referências bibliográficas consistente, discriminada ao final deste artigo.

Finalmente, para alcançar o objetivo central e responder a questão de pesquisa foram escolhi-

das, entre as bibliografias selecionadas, seis tecnologias que englobam os dados abertos vinculados, com foco no domínio da saúde e que visam agregar valor aos dados internos aos sistemas de informação, a saber:

a) *workbench*;
b) interfaces de navegação semântica;
c) *framework* no padrão *Linked Data*;
d) extratores do *Linked Data*;
e) *Linked Data Integration Framework* (LDIF), e
f) aplicativos *web* (especificamente o TripleMap).

Essas tecnologias serão explicadas em seção específica deste artigo.

Na seção a seguir são fundamentados os conceitos necessários para esclarecer o contexto de sistemas de informação (SI), em especial os clínicos (e os dados internos destes sistemas); a *web* semântica; o *Linked Data*; e o LOD. Não é intenção desse artigo exaurir todos estes conceitos, mas sim apresentá-los e fundamentá-los estabelecendo elos entre eles.

Informação Clínica e os Sistemas de Informação Clínicos

A Informação Clínica é aquela originada dos procedimentos relacionados ao tratamento da saúde de um indivíduo. São resultantes dos exames de laboratórios, procedimentos, entrevistas, internação hospitalar, pronto atendimento etc (VELDE; DEGOULET, 2003).

Alguns autores inserem os CIS como um subsistema dos Sistemas de Informação de Saúde de uma Comunidade (CHIS), os quais realizam a gestão direta dos pacientes (VELDE; DEGOULET, 2003). Contudo, sabe-se que esta nomenclatura se trata apenas de convenções. De fato, a necessidade de SI nas organizações tornou-se óbvia (LAUDON; LAUDON, 2011), o que não é diferente na área específica da saúde (SOCIEDADE..., 2013; SIGULEM, 1997).

A evolução da Informação Clínica se deu a partir dos anos 1960, quando os sistemas ainda eram denominados Sistemas de Informação de Hospital (HIS), abarcando informações médicas e administrativas. Os CHIS vieram em seguida e, a partir dos anos 1990, visavam à redução de custos, ajuda às instituições e comunidades médicas nas atividades diárias de tomada de decisão, bem como integração dos recursos e melhora da gestão do paciente. Os sistemas tiveram uma evolução natural sendo denominados CIS (SIGULEM, 1997).

Sistemas de Informação (SI) consistem num conjunto de partes que estão constantemente interagindo e se integrando, sempre com o propósito de atingirem objetivos e alcançar resultados. Os sistemas formam um todo unificado e nenhum sistema sozinho pode fornecer todas as informações que uma empresa necessita (LAUDON; LAUDON, 2011).

Para dar o aporte necessário aos dados dos SI necessita-se de um banco de dados (BD). O BD é um conjunto de dados inter-relacionados (LAUDON; LAUDON, 2011). O problema de interligar BD de SI distintos, apoiados em tecnologias heterogêneas, sempre foi um desafio e motivo de pesquisas, como pode ser visto em Hartig e Langegger (2010), Haase, Schimidt e Schwarte (2011), e Wylot *et al.* (2011).

Os dados e informações dos sistemas clínicos, ou seja, aqueles que são processados dentro das instituições médicas, hospitais, clínicas etc., caracterizam pela falta de interoperabilidade e são considerados internos ou fechados (VELDE; DEGOULET, 2003). Emerge, assim, a necessidade de interoperabilidade e integração dos dados (WORLD WIDE..., 2013). Primeiramente, os dados são produzidos de forma independente – o que acarreta heterogeneidade – e exigem intervenções voltadas para a construção de uma estrutura uniforme e integrada visando o compartilhamento (HARTIG; LANGEGGER, 2009).

Conceitos básicos relacionados à *web* semântica

Descrever, ainda que brevemente a *web* semântica é essencial para, posteriormente, elucidar o *Linked Data* e o LOD.

O *World Wide Web Consortium* (W3C) esclarece que a heterogeneidade semântica é um empecilho para alcançar a interoperabilidade, impedindo que os sistemas se comuniquem ou troquem informações (2013).

Dentre diversos conceitos, o próprio W3C define a interoperabilidade, de maneira ampla, como a capacidade de dois ou mais sistemas de interagir e trocar dados e informações, de acordo com métodos definidos e objetivando resultados esperados. Este Consórcio vem se dedicando no desenvolvimento de padrões para avançar rumo a excelência em termos de interoperabilidade. A partir do surgimento da *web* semântica a interoperabilidade na *web* está em processo de melhoria, pelo fato de oferecer a possibilidade de definir "expressividade" para os dados (WORLD WIDE..., 2013).

É notório que a *web* semântica é o resultado da aplicação de tecnologias de representação do conhecimento[2] em sistemas distribuídos em geral, com a finalidade de preencher o hiato de comunicação existente entre o ser humano e a máquina. No clássico artigo em 2001, "*The semantic web*", a web semântica é descrita como extensão da *web* atual, de textos e documentos, com o objetivo de desenvolver meios para que as máquinas possam servir aos humanos de maneira mais eficiente. Entretanto, é necessária a construção de instrumentos, no intuito de fornecer sentido lógico e semântico aos computadores (BERNERS-LEE; HENDLER; LASSILA, 2001).

Neste contexto, as ontologias cumprem um importante papel, pois conceituam formalmente um domínio com compromisso para o compartilhamento semântico. São instâncias (também denominados nós) representadas por relações que fazem sentido, formando mecanismos de controles terminológicos. Havendo uma ontologia entende-se que existe um consenso (GUARINO, 1995).

Adentrando no mundo da *web* semântica, Tim Berners-Lee partiu do princípio que todo recurso *web* (ou seja, qualquer conteúdo publicado na web) necessita de um *Uniform Resource Identifier* (URI) único[3]. Estas entidades, assim como são fundamentais para toda a *web*, formam a base da *web* semântica, pois nomeiam univocamente todo e qualquer recurso da *web* (BERNERS-LEE et al., 2006).

Em 1999 foi apresentada uma linguagem declarativa que se tornou um padrão posteriormente, chamada de *Resource Description Framework* (RDF), cuja base de escrita é em *eXtensible Markup Language* (XML), esta recomendada pelo consórcio W3C em 1998.

A representação do RDF é sempre feita em forma de uma sentença, ou tripla (sujeito – predicado – objeto). O sujeito tem o recurso (URI) o qual será escrita a sentença. O predicado ou propriedade, que também é um recurso (URI), representa o re-

Quadro 1 – Exemplos de triplas RDF

SUJEITO	PREDICADO	OBJETO
http://www.eci.ufmg.br/	rdf:hasName	"Escola de Ciência da Informação"
http://www.eci.ufmg.br/	rdf:isLocated	http://dbpedia.org/page/Minas_Gerais
http://dbpedia.org/page/Minas_Gerais	http://dbpedia-owl:type	http://dbpedia:States_of_Brazil

Fonte: elaborado pelos autores (2013).

[2] representar o conhecimento facilita a inferência e, a partir de elementos, é possível criar novos elementos (HALPIN; LAVERENKO, 2009).

[3] URI é uma *string* (cadeia) de caracteres, como por exemplos: http://www.google.com, que permite o acesso à página *web* do Google e http://en.wikipedia.org/wiki/Eiffel_Tower, que é a URI do recurso Torre Eiffel, na enciclopédia colaborativa Wikipedia.

lacionamento entre sujeito e um objeto. Já o objeto é o recurso ou um literal que se relaciona com o sujeito (WORLD WIDE..., 2103). O Quadro 1 apresenta exemplos de triplas RDF.

Os conjuntos de dados existentes na web semântica (*datasets*) são constituídos por triplas RDF, utilizando ligações (*links*) para os diversos conjuntos de dados participantes. O RDF pode ser considerado um modelo de dados que fornece uma semântica simplificada com representação para o tratamento de metadados (enquanto definição dos dados). No âmbito de tal definição de dados, o *Resource Description Framework Schema* (RDFs) é o esquema para declaração e definição de classes e tipos em RDF (WORLD WIDE..., 2013).

Contudo, pelo fato do RDF não fornecer subsídios necessários para uma expressividade exigida de uma ontologia para a *web* semântica, em 2004 o W3C formalizou a *Web Ontology Language* (OWL). A OWL descreve os aspectos semânticos dos termos utilizados e seus relacionamentos, favorecendo uma representação mais abrangente do RDF tendo em vista a interoperabilidade. Nos anos seguintes à sua formalização, surgiram novas versões da OWL, como a 2.0 em 2009, que provê a escalabilidade necessária para sua evolução (WORLD WIDE..., 2013).

Complementando estes elementos da *web* semântica, o *Simple Protocol and RDF Query Language* (SPARQL) foi recomendado em 2008 pelo W3C. É uma linguagem (e também um protocolo) de consulta RDF para expressar *queries* (consultas realizadas à base de dados) em diversas fontes de dados armazenados nativamente em RDF. Já o SPARQL *endpoint* é um serviço para implantação do SPARQL, o qual permite consultas a uma base de dados em RDF. Para que as aplicações acessem os dados do LOD, é necessário efetivar consultas SPARQL em um SPARQL *endpoint* (WORLD WIDE..., 2013).

Uma vez apresentados os conceitos da *web* semântica, para a continuidade deste estudo, torna-se necessário elucidar os conceitos do *Linked Data* e do LOD, a seguir.

O *Linked Data*, o LOD e seus fundamentos

O *Linked Data* descreve um conjunto de práticas para publicar e conectar dados estruturados na *web*. Agrega os mesmos princípios básicos da *web* propostos por Berners-Lee: simplicidade, design modular e descentralização (BERNERS-LEE, 2001).

Berners-Lee *et al*. (2006) elaboraram os princípios do *Linked Data*, esclarecendo que os dados em RDF nesta estrutura devem ser:

a) abertos (e não proprietários), pois devem ser acessados por meio de uma ilimitada variedade de aplicações;
b) modulares, uma vez que não necessitam de planejamento prévio para combinar com outros dados;
c) escaláveis, pois uma vez que já existam dados em RDF, a adição de novos dados tende a ser facilitada.

O W3C, desde o surgimento dos princípios do *Linked Data*, oferece o suporte necessário ao projeto de dados abertos vinculados (*Linked Open Data* – LOD) e, com isto, impulsiona a produção de dados na *web*.

Além do crescimento significativo dos *datasets* do LOD, o vocabulário heterogêneo dos dados, aliada à fragmentação natural no ambiente *web*, dificulta seu consumo e reutilização. Mecanismos cada vez mais eficientes são criados a fim de permitir a utilização por qualquer interessado (BIZER; HEATH; BERNERS-LEE, 2009).

Um destes mecanismos é a apresentação dos dados na forma de nuvem, cuja necessidade de criação advém da facilidade de se ter um diagrama completo, agregando todos os *datasets* com dados em forma de triplas RDF, assim como suas relações com outros *datasets*.

O Diagrama de Nuvem do *Linked Open Data*

O diagrama de nuvem é mantido por Richard Cyganiak, do *Digital Enterprise Research Institute* (DERI[4]), da Universidade Nacional da Irlanda/ Galway, e Jentzsch Anja, da Universidade Freie[5], Berlin. Este diagrama se destina a todos que desejam participar do projeto do LOD, acessando, consumindo ou publicando as triplas RDF.

A imagem da nuvem mostra os conjuntos de dados que são publicados e como estão interligados com outros conjuntos de dados. Outra característica importante é a funcionalidade de oferecer cliques nos *datasets*, possibilitando a navegação. Ao ser clicado, o *dataset* abre todas as suas características e as triplas RDF que possui[6].

De Maio/2007 a Setembro/2011 houve onze novas "nuvens" publicadas. Abaixo a tabela com a lista das datas de publicação e quantidade de *datasets*.

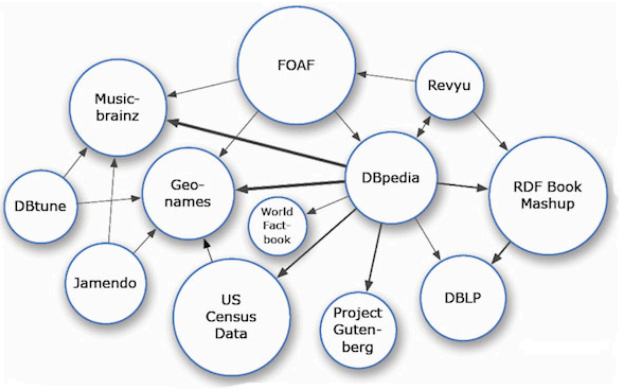
Tabela 1 – Publicação dos *datasets*

Quantidade de nuvens	Data Publicação	Número de datasets
1	01/05/2007	12
2	08/10/2007	25
3	07/11/2007	28
4	28/02/2008	32
5	31/03/2008	34
6	18/09/2008	45
7	05/03/2009	89
8	27/03/2009	93
9	14/07/2009	95
10	22/09/2010	203
11	19/09/2011	295

Fonte: LINKED Data (2013), adaptado pelos autores.

Chama-se a atenção para o crescimento acentuado de Julho/2009 para Setembro/2010. Na primeira publicação de tais conjuntos (Maio de 2007) existiam doze 12 *datasets* disponíveis (Figura 1). Em Setembro de 2011, todavia, este número já alcançava 295 *datasets*.

Figura 1 – Diagrama de Nuvem LOD – Maio/2007



Fonte: LINKED Data, 2013.

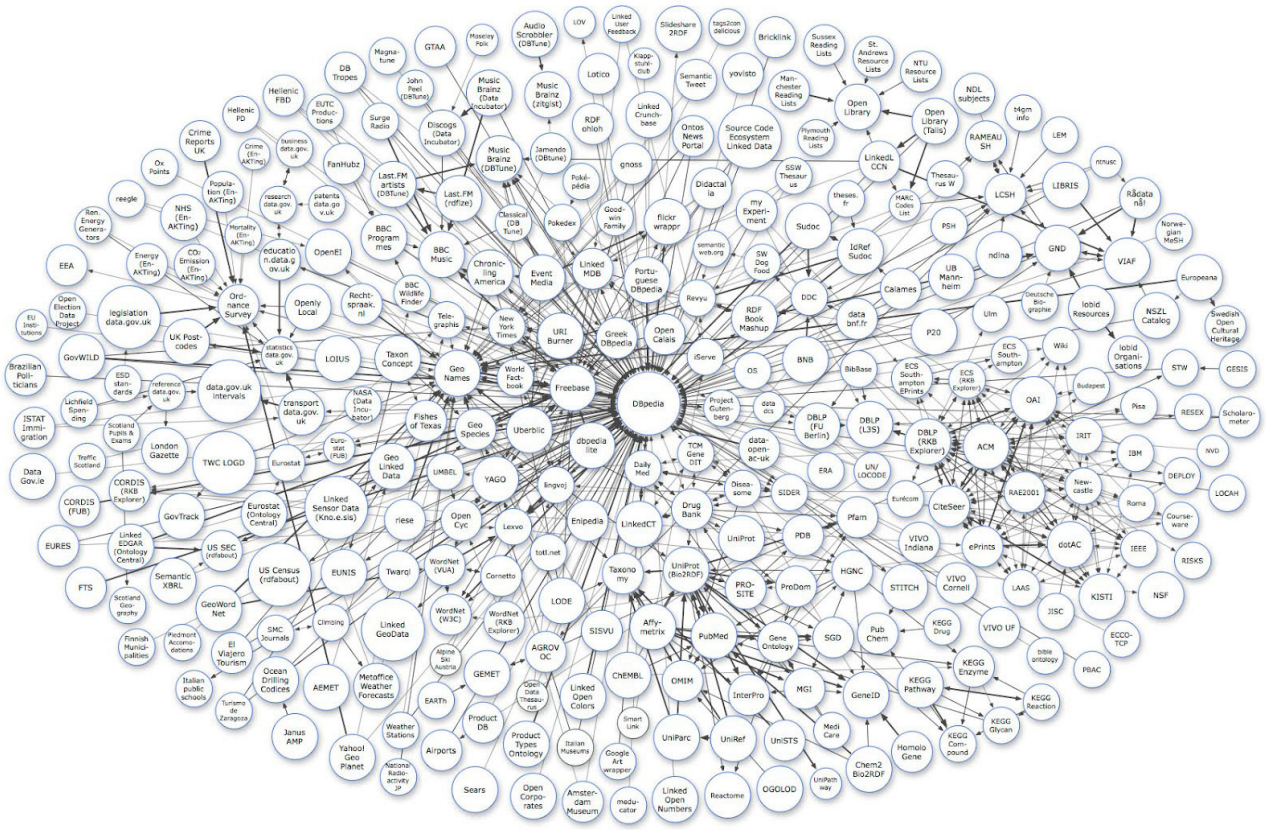
O tamanho dos círculos corresponde ao número de triplas RDF que possuem. As setas bidirecionais (*BDPedia – Revyu*) indicam que os *links* estão em ambos os conjuntos de dados. Já a seta em uma única direção (de DBPedia para US Census Data), indica que o *dataset* de origem contém triplas RDF que são consumidas no *dataset* de destino.

A intenção em apresentar a Figura 2, representando a nuvem de dados mais atual (LINKED..., 2013), repousa no motivo de sua completude. Pode ser verificado ao centro o *DBPedia*, *dataset* do Wikipedia, que possui o maior número de conexões, pois dos dados gerais do Wikipedia são consumidos por diversos domínios. Chama-se a atenção para o lado direito do *dataset DBPedia*, onde encontram-se diversos *datasets* acadêmicos, como o ACM. Do lado esquerdo, encontra-se a área de entretenimento, com o BBC *Music*, dentre outros.

Cabe destacar o grupo de *datasets* relacionados ao projeto do LOD, o *Linked Open Drug Data* (LODD[7]), que está situado abaixo do *BDPedia*. No intuito de facilitar a conexão de um grupo de dados específico, a saúde, este esforço trabalha com um conjunto de tecnologias e convenções

[4] http://www.deri.ie/.
[5] http://www.fu-berlin.de/en.
[6] esta funcionalidade pode ser visualizada em: http://lod-cloud.net/versions/2011-09-19/lod-cloud.html.
[7] http://www.w3.org/wiki/HCLSIG/LODD.

Figura 2 – LOD Cloud Diagram – Setembro/2011



Fonte: BIZER; JENTZSCH; CYGANIAK (2011).

facilitando a conexão dos dados. Outro fator motivador para o LODD é a significativa quantidade de informações sobre drogas disponíveis na *web* uma vez em que há, desde sua concepção, o esforço do *Linked Open Data* na área médica (Saúde e Ciências da Vida).

É fato que apenas consumir – termo utilizado no projeto do LOD para utilizar as triplas RDF do LOD – não basta. É preciso publicá-las para que o esforço do projeto em agregar mais conteúdo seja alcançado. Neste sentido, Heath e Bizer (2011) propõem uma completa fonte de referência para os que desejam se tornar um dos publicadores de *datasets* do LOD.

Publicando e consumindo RDFs no LOD

Heath e Bizer (2011) esclarecem que basicamente são dois tipos de aplicações que consomem dados do LOD:

a) aplicações genéricas, que fazem o uso dos dados do LOD em qualquer domínio;
b) aplicações que focam em um domínio específico.

O consumo (utilização) dos dados do LOD necessita, incialmente, de conceitos como URI, RDF, RDFs e SPARQL, apresentados anteriormente. O acesso pode ser feito por meio do conhecimento prévio do vocabulário das fontes de dados e da sintaxe das consultas SPARQL, submetidas nos SPARQL *endpoints*. Esta abordagem é denominada tradicional, como explicam Hartig e Langegger (2010), e os dados são materializados em um *data warehouse*, que é um local de grande capacidade onde podem ser armazenados dados para consultas posteriores.

Em outra abordagem, denominada federação de consultas, prevê "um mediador transparente que decompõe a consultas em subconsultas e encaminha as subconsultas a múltiplos serviços de consulta distribuídos" (PINHEIRO, 2011, p. 17-18).

O acesso aos dados é realizado por meio de um vocabulário padrão especificado na ontologia de domínio, as quais contêm termos essenciais de um domínio específico do conhecimento, diferentemente das ontologias fundacionais (GUARINO, 1995). Depois de obtidos os resultados, os dados são integrados e a resposta final é entregue ao usuário. Existem algumas vantagens nesta abordagem como, por exemplo, a economia de tempo e de espaço adicional para materialização de dados, e a obtenção de dados totalmente atualizados, uma vez que a extração ocorre no momento em que a consulta é requisitada (PINHEIRO, 2011). Trabalhos como o de Pinheiro (2011) vêm sendo desenvolvidos no âmbito de aperfeiçoar estas consultas.

Torna-se importante contextualizar o conceito de objetos não referenciados, ou seja, para os quais não se conhecem, a princípio, suas respectivas URIs. Neste caso, quando se obtêm tais informações, o resultado é uma descrição RDF do recurso identificado, resultando na acumulação de benefícios tais como a criação de *links* RDF (ou URI-*links*) entre dados de diferentes fontes de dados (ALLEMANG; HENDLER, 2011).

Para que os dados sejam publicados em *Linked Data*, Heath e Bizer (2011) indicam um conjunto de princípios deverá ser seguido[8]. Alguns deles são:

a) usar URIs únicas para todos os recursos e evitar alterá-las para não haver quebra de *links* já estabelecidos;
b) as URIs devem ser passíveis de procura e recuperação (URIs não referenciáveis);
c) fornecer dados no formato RDF;
d) conter pelo menos 1.000 triplas;
e) ser conectado com pelo menos 50 *links* RDF, uni ou bidirecionais, para um conjunto de dados que já está no diagrama da nuvem (FIGURA2);
f) ser acessado por SPARQL *endpoint*, ou *browsers* do *Linked Data*, como o "Tabulator"[9] (BERNERS-LEE *et al.*, 2006) que permitem percorrer a *web* de dados.

Com relação ao armazenamento de dados no padrão necessário do *Linked Data*, Heath e Bizer (2011) indicam as seguintes formas:

a) utilizar uma *Application Programming Interface* (APIs) para as triplas RDFs nativas. Exemplos: Sesame[10] e Jena[11];
b) fornecer *wrappers* (tradutores) para banco de dados relacionais. Neste caso, Bizer e Cyganiak (2006) propõem esta funcionalidade. Um exemplo é o "Virtuoso"[12];
c) fazer a triplificação (disponibilização de dados em triplas RDFs) de fontes de dados relacionais. Os exemplos são o Jena SDB[13] e Jena TDB[14].

Finalizando a fundamentação dos conceitos apresenta-se, a seguir, a interligação do LOD com os dados internos dos SI.

Finalizando a fundamentação dos conceitos apresenta-se, a seguir, a interligação do LOD com os dados internos dos SI.

Os *datasets* do LOD como valor agregado aos dados internos de Sistemas de Informação Clínicos

Há uma diversidade no conceito "valor agregado", inclusive quando se privilegia o cliente no centro do processo e quando este estabelece a sua satisfação com um produto, um serviço ou um

[8] a lista completa dos princípios encontra-se no link <http://www4.wiwiss.fu-berlin.de/bizer/pub/linkeddatatutorial/>. Acesso em: 11 abr. 2013.
[9] http://www.w3.org/2005/ajar/tab.
[10] http://www.openrdf.org/.
[11] http://jena.sourceforge.net/.
[12] http://virtuoso.openlinksw.com/.
[13] http://openjena.org/TDB/.
[14] http://openjena.org/TDB/.

sistema de informação. O foco no produto significa, normalmente, a incorporação de tecnologias e uma sofisticação do mesmo. Contudo, agregar valor aos dados de sistemas de informações significa a possibilidade de não tê-los em ambientes isolados, mas sim de forma interoperável, transformando-os em informações passíveis de análises estratégicas integradas para uma tomada de decisão precisa (LAUDON; LAUDON, 2011).

Velde e Degoulet (2003) explicam que, em geral, se tende a focalizar os esforços na quantidade de dados no contexto da informação clínica, muitas vezes não proporcionando qualquer tipo de análise, ou capacidade de rastreabilidade, ou mesmo de historicidade dos dados. O mais importante, de acordo com estes autores, é que seja viável tal análise pelos sistemas de informação.

É impossível imaginar a possibilidade de agregar valor aos dados de informação clínica sem especificar e consumir alguns *datasets* do LOD voltados para o domínio da saúde. Existem 41 (de um total de 295[15]) *datasets* com mais de três bilhões de triplas na área de ciências da vida (*life sciences*) (BIZER, JENTZSCH; CYGANIAK, 2011). Acima deste número encontram-se apenas 87 *datasets* de publicações acadêmicas e 49 *datasets* de dados governamentais. Alguns destes *datasets* no domínio das ciências da saúde são listados abaixo (LINKED..., 2013):

a) *DailyMed*: publicado pela Biblioteca Nacional de Medicina fornece informações de qualidade sobre drogas comercializadas;
b) *Diseasome*: rede pública de mais de 4.300 doenças e genes ligados a distúrbios;
c) *DrugBank*: repositório de quase 5.000 moléculas e informações sobre drogas;
d) *Gene Ontology* (GO): iniciativa da área de bioinformática voltada à unificação da representação dos atributos dos genes de todas as espécies;
e) *InterPro*: banco de dados de famílias de proteínas, advogando possuir as mais novas proteínas;
f) SIDER: contém informações sobre drogas comercializadas e seus efeitos colaterais extraídas de documentos públicos e de bulas;
g) STITCH: contém informações sobre produtos químicos e proteínas;
h) *TaxonConcept*: considerando-se que as espécies são conhecidas por muitos nomes diferentes. Esta base de conhecimento agrega URIs para distintos conceitos de espécies.

Uma possibilidade de ligação de recursos (dados) internos e externos pode ser feita por meio da URI (HEATH; BIZER, 2011). Estes autores exemplificam esta ligação utilizando chaves primárias em banco de dados internos, tal como o *International Standard Book Number* (ISBN) como ligação com dados do LOD. A seguir apresentam-se tecnologias que podem ser utilizadas para agregar valor aos dados internos de SI na área da saúde.

Tecnologias voltadas aos *datasets* LOD: valor agregado potencial para Sistemas de Informação Clínicos

As tecnologias foram selecionadas nas referências obtidas quando da pesquisa documental e têm como base os princípios do *Linked Data* e os dados do LOD. Inicia-se a descrição com tecnologias genéricas e finaliza-se com aquelas específicas para o domínio da saúde, as quais podem ser utilizadas nos sistemas de informação clínicos. São elas:

a) *workbench*: conjunto integrado de ferramentas para apoiar o processo computacional no intuito de diminuir a barreira de entrada para a dimensão do *Linked Data*;
b) interfaces de navegação com o uso de tecnologias semânticas;
c) *framework*, como uma estrutura (ou arcabouço) conceitual no intuito de facilitar a agregação, baseado em mediador para a integração de dados no padrão *Linked Data*;

[15] conforme a publicação da nuvem de *datasets* em 19 set. 2011.

d) extratores do *Linked Data*;
e) *Linked Data Integration Framework*;
f) aplicativos *web*.

Workbenchs para ligação dos dados no padrão *Linked Data*

Haase, Schmidt e Schwarte (2011) propõem um *workbench* no intuito de diminuir a barreira de entrada para o mundo do *Linked Data*. O apoio à descoberta e exploração de fontes de dados facilita a integração e processamento de dados abertos vinculados. Fontes de dados remotas podem ser virtualmente integradas através de uma camada de federação, e o desenvolvimento de uma interfase de usuário *self-service* igualmente facilita a transição. O uso é baseado em uma *wiki* semântica[16], combinados com um conjunto robusto de *widgets*[17] para interação com os dados. Vale ressaltar que esta abordagem de integração fica transparente para o usuário. Não existe a preocupação com aspectos da distribuição física dos dados ou protocolos de acesso, pois detalhes da integração ficam ocultos em tempo de execução. Dados locais e fontes virtualmente integradas podem ser consultados de forma integrada.

Uma empresa, a Fluid Operations[18], oferece o "*Information Workbench – for a world where all data is Linked*", que é uma plataforma *web* aberta para soluções de *Linked Data* para organizações. Dados de diferentes fontes são integrados e conectados, utilizando uma camada de dados do *Linked Data* por sobre o conteúdo, facilitando o acesso semântico e a busca inteligente. A Fluid Operations definiu sua plataforma com base nos estudos de Haase, Schimidt e Schwarte (2011).

Interfaces de navegação com o uso de tecnologias semânticas

Em uma importante abordagem de ligação de dados internos, Passant *et al.* (2010) propõem interfaces de navegação e *mashups*[19] semânticos. A novidade desta aplicação está na reutilização de dados RDF do *GeoNames*[20] que fornece um *mashup* semântico da combinação de fonte dados externos e internos. Os dados internos são combinados com dados do *GeoNames* a partir das coordenadas de localização, permitindo uma navegação avançada dos recursos. A representação visual é outro ponto alto desta aplicação. Os autores acreditam que *mashups* semânticos podem ser parte significativa do futuro das aplicações Enterprise 2.0, termo que indica uma linha de evolução das organizações voltadas para o conhecimento que utilizam ferramentas da *web* 2.0 (cooperação e colaboração) e da *web* 3.0 (semântica). O uso do *Linked Data* é fundamental para este sucesso, na medida em que as empresas se beneficiam de informações públicas a custo zero. Passant *et al.* (2010) complementam a importância desta ligação inclusive com dados legados das organizações.

Framework baseado em mediador para a integração de dados no padrão *Linked Data*

Pinheiro (2011) apresenta um cenário com a necessidade de integração a partir de múltiplas fontes de dados públicas. O autor propõe um esquema mínimo de fontes de dados da área médica com respectivas ligações para a integração no padrão *Linked Data*. Os *datasets* escolhidos pelo autor incluem informações sobre doenças (*Diseasome*), drogas (*DrugBank*), bulas de drogas (*DailyMed*), medicamentos e efeitos diversos (*Sider*), assim como do *DBPedia*, que interliga

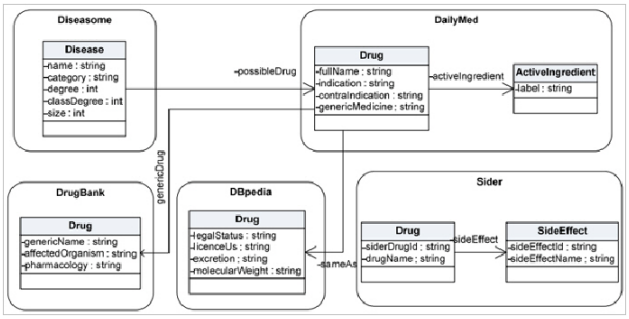
[16] conceito que implementa tecnologias da *web* semântica nas ferramentas *wikis*.
[17] pequenos aplicativos que utilizam interface gráfica e que funcionam nos sistemas web. Incluem janelas, botões, menus, ícones, barras de rolagem, dentre outros.
[18] http://www.fluidops.com.
[19] *sites* ou aplicações *webs* que usam informações de mais de uma fonte no intuito de criar novos serviços.
[20] *dataset* do LOD com informações sobre mais de 6 milhões de lugares e características geográficas.

praticamente todos os domínios do LOD como uma fonte de dados de temas variados.

Esta pesquisa apresenta um *framework* baseado em mediador para a integração de dados no padrão *Linked Data*, acessíveis via SPARQL *endpoint*, que é um serviço para implantação de consultas no modelo de dados RDF. O esquema de mediação é representado por uma ontologia de domínio, que fornece um vocabulário compartilhado (PINHEIRO, 2011).

Além disto, é proposto um método para o processamento distribuído de consultas SPARQL. O autor fornece uma "solução não intrusiva e de fácil utilização para processamento de consultas em um ambiente mediado no contexto de dados publicados no padrão de *Linked Data*" (PINHEIRO, 2011, p. 143). A Figura 3 apresenta os *datasets* do LOD da área médica que foram escolhidos pelo autor, assim como os atributos e relacionamentos.

Figura 3 – Interligação de fonte de dados do LOD no domínio médico no padrão Linked Data



Fonte: PINHEIRO (2011, p. 20).

Extratores do *Linked Data*

A *web* dos dados oferece a ideia concreta de que mais e mais dados estão interligados. Rizzo e Troncy (2011) consideram que, para almejar um foco estruturado dos dados, existe a necessidade de proporcionar anotações mais estruturadas aos documentos, utilizando vocabulários comuns ou ontologias. Textos semiestruturados, como os de natureza científica, médica ou artigos de notícias, têm uma maior possibilidade de serem semanticamente anotados. Entidades extratoras desempenham um papel fundamental para a retirada de informações estruturadas, identificando suas características (entidades) e ligando-as a outros recursos da *web* por meio de inferências.

Os autores desenvolvem um trabalho que agregam valor aos dados, onde avaliam os extratores mais populares do *Linked Data*, como *DBPedia Spotlight*, *Extractiv*, *OpenCalais*, *Alche-MyAPI* e *Zemanta*. O resultado da pesquisa é um *framework* com avaliação realizada pelo homem, atribuindo um valor a detecção da entidade, tipo de entidade e desambiguação (RIZZO; TRONCY, 2011).

Heath e Bizer (2011), ao discutirem os extratores para o *Linked Data* reforçam que, onde a entrada de dados é textual e há recursos naturais de linguagem – como, por exemplo, uma série de notícias ou relatórios de negócios – é possível "passar" estes documentos através de extratores de entidades para o *Linked Data*. Publicar estas anotações junto dos documentos melhora a tarefa de recuperação da informação e permite aplicativos usarem as fontes do *Linked Data* referenciadas, como um pano de fundo para mostrar as informações sobre as páginas, além de oferecer a navegação facetada.

Linked Data Integration Framework

Schultz *et al.* (2011) desenvolveram um *framework* para a construção de aplicações de dados do LOD. O *Linked Data Integration Framework* (LDIF) traduz dados vinculados heterogêneos da *web* em uma representação mais limpa voltada para o uso local, ou seja, para os dados internos dos SI, mantendo a procedência dos dados. Fornece uma linguagem de mapeamento expressiva para traduzir os dados de vocabulários diferentes com vistas ao uso local. Inclui também um componente de resolução de identidade que descobre a URI baseado nos dados de entrada.

O estudo de caso destes autores foi, justamente, o domínio da ciência da vida, presente em dois datasets do LOD, a saber:

a) Kegg Gened: uma coleção de catálogos de genes gerada a partir de recursos públicos;
b) *UniProt*: um conjunto de dados contendo sequências de proteínas, genes e funções.

Os autores pretendem estender as funcionalidades do LDIF, contemplando, principalmente:

a) a distribuição dos processos em *cluster*, cujo conceito de conjunto de máquinas distribuídas, trabalhando como se fossem uma única máquina, permite uma escala de processamento em uma grande de quantidade de dados;
b) a utilização de *crawlers*[21] para *Linked Data* e SPARQL *endpoint*;
c) o uso de recurso adicional para avaliação da qualidade dos dados do *Linked Data*.

Aplicativos *web*

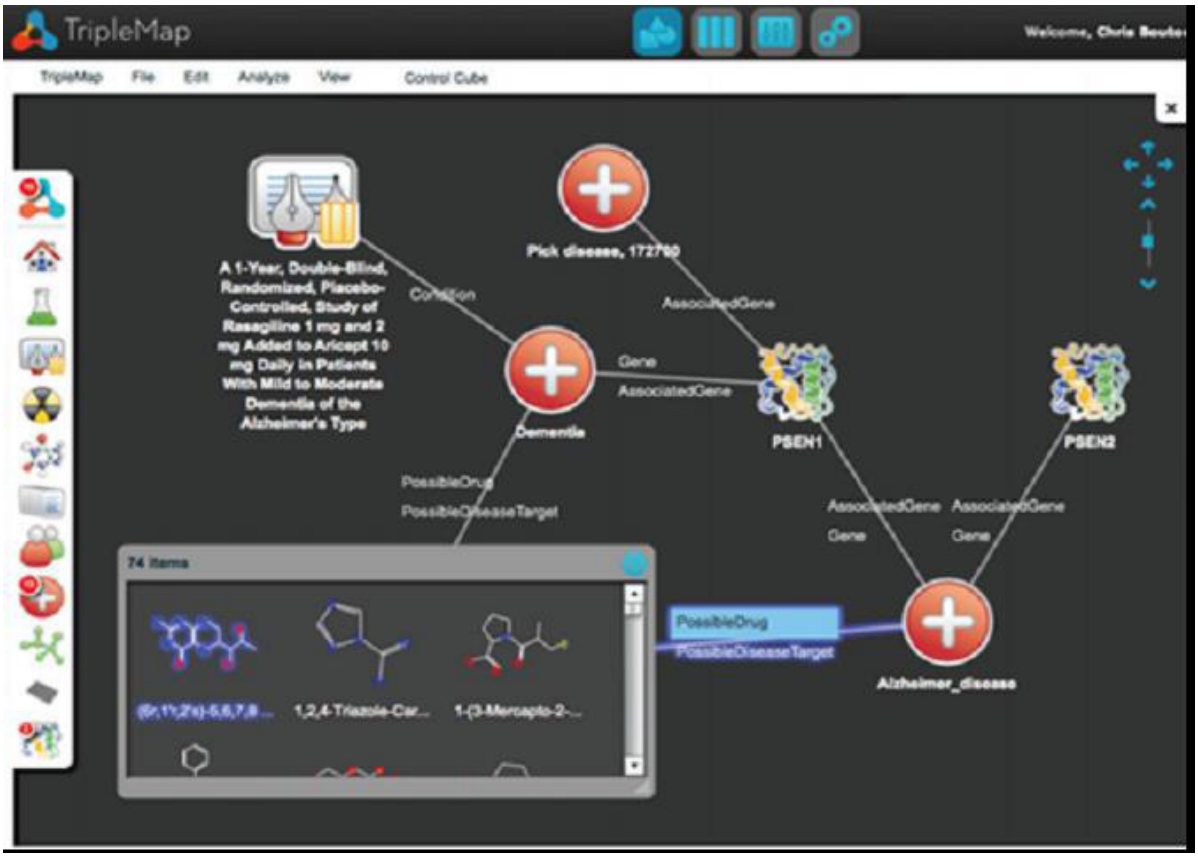
Samwald *et al.* (2011) apresentam o *TripleMap*, um aplicativo para *web* que fornece uma interface rica, dinâmica e integrada, para conjunto de dados RDF do LOD. Este permite a escolha de qual o domínio se deseja trabalhar como, por exemplo, o domínio da saúde, que resulta na localização d *Linked Open Drug Data* (LODD), o qual está composto por recursos *web* relativos à doenças, drogas, ensaios de pesquisa etc. (Figura 4).

Na Figura 4 estão representadas as entidades e suas associações, as quais podem ser arrastadas para dentro da tela, possibilitando uma navegação automática.

Considerações finais

No decorrer deste estudo apresentou-se uma discussão dos sistemas de informação, em especial os clínicos (CIS), existentes nas instituições mé-

Figura 4 – Tela do TripleMap com datasets do LODD



Fonte: SAMWALD et al. (2011).

[21] termo que significa rastreadores na *web*, na forma de robôs, com o propósito de indexação automática ou atualização de conteúdo.

dicas e utilizados na área da saúde para a gestão de pacientes. Identificou-se que tais sistemas carecem de interoperabilidade, sendo considerados sistemas internos ou fechados.

A efetivação da interoperabilidade dos sistemas se dá na medida em que, na *web* semântica, a expressividade dos dados pode ser alcançada utilizando os conceitos de URI, RDF, RDFs, OWL e SPARQL. Para fundamentar o conceito da *web* semântica – considerada a *web* dos dados, em contrapartida a *web* dos textos e documentos – utilizaram-se autores que participaram da sua proposição, tais como Tim Berners-Lee e James Hendler.

Ao elucidar os fundamentos dos padrões do Linked Data e do projeto *Linked Open Data* (LOD), recorreu-se a autores seminais, tais como Christian Bizer, Richard Cyganiak e Jentzsch Anja. Estes dois últimos mantém a nuvem dos *datasets* do LOD, detalhada na fundamentação dos conceitos. Utilizaram-se os *sites* de referência e que encarregam da proposição, validação e acompanhamento do LOD, como o consórcio W3C e o *Linked Data* (www.linkeddata.org). Introduziram-se, também, elementos voltados ao estudo de publicação e consumo do LOD.

No que diz respeito às nuvens LOD, mesmo quando há a possibilidade de se utilizar mais de três bilhões de triplas RDF (isto apenas nos 41 *datasets* do LOD no domínio da saúde[22]), já é possível agregar valor aos dados internos de sistemas de informação na área clínica. Esta possibilidade permitirá análises integradas que extrapolarão o uso atual de dados internos a tais sistemas.

O caráter exploratório deste estudo conduziu à necessidade de examinar tecnologias que utilizem os *datasets* do LOD como forma de agregar valor aos sistemas de informação, em especial os clínicos. Tais tecnologias englobam o *workbench*, as interfaces de navegação semântica, o *framework* no padrão *Linked Data*, os extratores do *Linked Data*, o *Linked Data Integration Framework* (LDIF) e os aplicativos *web*, tais como o *TripleMap*. Os pontos em comum dessas tecnologias situam-se na publicação, extração, interligação e consumo dos dados do LOD.

Este artigo alcança seu objetivo que foi apresentar o LOD, sua evolução e nuvem, e a possibilidade de agregar valor aos dados dos CIS. Contudo, é apenas o início de um estudo em um vasto campo de pesquisa, utilizando principalmente os *datasets* do LOD no domínio da saúde. Identificou-se estudos embrionários de diversos domínios, espalhados pelas principais universidades do mundo, principalmente no DERI e na Universidade de Freie, citadas neste artigo, cujas referências contemplaram o estado da arte.

A continuidade deste estudo documental se dará na medida em que se faça uma aplicação prática do mesmo, ou seja, utilizando uma ou mais tecnologias aqui apresentadas, será factível comprovar o valor que se pode agregar ao utilizar as triplas RDF do LOD no domínio da saúde, aos dados internos dos sistemas de informação clínicos.

[22] dados de setembro/2011.

Referências

ALLEMANG, D.; HENDLER, J. **Semantic web for the working ontologist**: effective modeling in RDFS as OWL. 2. ed. Waltham, USA: Elsevier, 2011.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, p. 34-43, May 2001. Disponível em: <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>. Acesso em: 22 jun. 2013.

BERNERS-LEE, T.; CHEN, Y.; CHILTON, L.; CONNOLLY, D.; DHANARAJ, R.; HOLLENBACH, J.; LERER, A.; SHEETS, D. Tabulator: exploring and analyzing linked data on the semantic web. In: INTERNATIONAL SEMANTIC WEB USER INTERACTION, 3. 2006. **Proceedings…** Disponível em: <http://swui.semanticweb.org/swui06/papers/Berners-Lee/Berners-Lee.pdf>. Acesso em: 4 jun. 2013.

BIZER, C.; CYGANIAK, R. Publishing relational databases on the Web as SPARQL- Endpoints. In: INTERNATIONAL WORLD WIDE WEB CONFERENCE, 15., 2006. **Proceedings…** Edinburg, 2006.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data: the story so far. **International Journal on Semantic Web and Information Systems**, v. 5, n. 3, p. 1-22, 2009. Disponível em: < http://eprints.soton.ac.uk/271285/1/bizer-heath-berners-lee-ijswis-linked-data.pdf>. Acesso em: 4 jun. 2013.

BIZER, C.; JENTZSCH, A.; CYGANIAK, R. **State of the LOD Cloud**. 2011. Disponível em: <http://lod-cloud.net/state/>. Acesso em: 4 jun. 2013.

GUARINO, N. Formal ontology, conceptual analysis and knowledge Representation. **International Journal of Human-Computer Studies**, v. 43, n. 5/6, p. 625-640, 1995. Disponível em: <http://nemo.nic.uoregon.edu/wiki/images/7/79/Guarino_IJHCS1995_Formal_Onto_conceptual_analysis.pdf>. Acesso em: 4 jun. 2013.

HAASE, P.; SCHMIDT, M.; SCHWARTE, A. The information workbench as a self-service platform for linked data applications. In: INTERNATIONAL WORKSHOP ON CONSUMING LINKED DATA, 2., 2011. **Proceedings…** Bonn, Germany: [S.n.], 2001. Disponível em: <http://ceur-ws.org/Vol-782/HaaseEtAl_COLD2011.pdf>. Acesso em: 4 jun. 2013.

HALPIN, H.; LAVERENKO, V. Relevance feedback between hypertext and semantic search. In: SEMANTIC SEARCH WORKSHOP AT THE WORLD WIDE WEB CONFERENCE, 18., 2009. **Proceedings…** Madrid: [S.n.], 2009. Disponível em:<http://km.aifb.kit.edu/ws/semsearch09/semse2009_27.pdf>. Acesso em: 4 jun. 2013.

HARTIG, O.; BIZER, C.; FREYTAG, J.-C. Executing SPARQL queries over the Web of Linked Data. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 8., Chantilly, VA, USA, 2009. **Proceedings…** Berlin: Springer Berlin Heidelberg, p. 293-309. Disponível em: <http://dx.doi.org/10.1007/978-3-642-04930-9_19>. Acesso em: 4 jun. 2013.

HARTIG, O.; LANGEGGER. A database perspective on consuming linked data on the Web. **Datenbank-Spektrum**, v. 10, n. 2, p. 57-66, 2010. Disponível em: <http://www.vldb.informatik.hu-berlin.de/~hartig/files/Hartig_QueryingLD_DBSpektrum_Preprint.pdf>. Acesso em: 4 jun. 2013.

HEATH, T.; BIZER, C. **Linked Data**: Envolving the web into a global data space. California, USA: Morgan & Claypool, 2011.

LAUDON, K. C.; LAUDON J. P. **Sistemas de informação gerenciais**: administrando a empresa digital. 7. ed. São Paulo: Prentice Hall, 2011.

LINKED data. Disponível em: <www.linkeddata.org>. Acesso em: 4 jun. 2013.

PASSANT, A.; LAUBLET, P.; BRESLIN, J. G.; DECKER, S. Enhancing enterprise 2.0 ecosystems using semantic web and linked data technologies: The SemSLATES approach. In: WOOD, D. (Org.). **Linking Enterprise Data**. New York: Springer, 2010.

PINHEIRO, J. C. **Processamento de consulta em um framework baseado em mediador para integração de dados no padrão de Linked Data**. 2011. Tese (Doutorado em Ciência da Computação), Universidade Federal do Ceará, 2011. 156 p.

RIZZO, G.; TRONCY, R. NERD: A framework for evaluating named entity recognition tools in the Web of data. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 10., 2011 **Proceedings…** Bonn, Germany: [S.n], 2011. Disponível em: <http://www.eurecom.fr/fr/publication/3515/download/mm-publi-3515.pdf>. Acesso em: 4 jun. 2013.

SAMWALD, M.; JENTZSCH, A.; BOUTON, C.; KALLESØE, C. S.; WILLIGHAGEN, E.; HAJAGOS, J.; MARSHALL, M. S.; PRUD'HOMMEAUX, E.; HASSANZADEH, O.; PICHLER, E.; STEPHENS, S. Linked open drug data for pharmaceutical research and development. **Journal of Cheminformatics**,

v. 3, n. 19, 2011. Disponível em: <http://dx.doi.org/10.1186/1758-2946-3-19>. Acesso em: 4 jun. 2013.

SOCIEDADE Brasileira de Informática em Saúde. Disponível em: <http://www.sbis.org.br>. Acesso em: 4 jun. 2013.

SCHULTZ, A.; MATTEINE, A.; ISELE, R.; BIZER, C.; BECKER, C. LDIF: Linked Data Integration Framework. In: INTERNATIONAL WORKSHOP ON CONSUMING LINKED DATA, 2., 2011, Bonn. **Proceedings…** Aachen, Germany: CEUR, 2011. Disponível em:<http://ceur-ws.org/Vol-782/SchultzEtAl_COLD2011.pdf>. Acesso em: 4 jun. 2013.

SIGULEM, D. **Um novo paradigma de aprendizado na prática médica da UNIFESP/EPM**. Tese (Livre Docência do Centro de Informática em Saúde – CIS – EPM), Escola Paulista de Medicina, 1997.

VELDE, R. V. de; DEGOULET, P. **Clinical Information Systems**: a Component-Based Approach. New York: Springer-Verlag, 2003.

WORLD WIDE WEB CONSORTIUM. Disponível em: <www.w3.org>. Acesso em: 4 jun. 2013.

WYLOT, M.; PONT, J.; WISNIEWSKI, M.; CUDRÉ-MAUROUX, P. dipLODocus[RDF] short and long-tail RDF analytics for massive webs of data. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 10., 2011. **Proceedings...** Berlin: [S.n.], 2011. p. 778-793.

Linked Open Data as a way to add value to clinical information

Abstract

Introduction: Linked Open Data (LOD) has been a constant subject in the main semantic web conferences and journals around the world. Some studies prove that the consumption of these data is potentially important to improve the data quality of the information systems in many areas. The clinical information is one of these areas and the LOD initiative has been trying to standardize its publication, becoming the link of more efficient datasets. However, adding value to internal data of information systems is challenging. The aim of this paper is to present LOD in order to verify the possibility of adding value to the clinical information systems' data. Method: Based on the researchers who proposed the semantic web those who follow LOD's evolution, a literature revision of the main concepts was made. A bibliographic research was developed obtaining a reference based on books and papers of well known authors. Results: Were presented six technologies that utilize the Linked Data pattern and propose the addition of value to the system information's data, especially the clinical ones. The common points of these six technologies are situated in the publication, extraction, link and consumption of LOD. Conclusions: It is possible to add value to internal data of the clinical systems, even when more than three billion LOD triples are available to be consumed.

Keywords

Semantic Web. Information systems. Linked open data.

Recebido em 5 de abril de 2013
Aceito em 3 de junho de 2013

Sobre os autores:

Fernando Hadad Zaidan
Bacharel em Ciência da Computação - FUMEC, Mestre em Administração - FUMEC, Doutorando em Ciência da Informação - UFMG.
fhzaidan@gmail.com

Marcello Peixoto Bax
Bacharel em Ciência da Computação - PUC-MG, Mestre DEA en Informatique et Matematique - Université d Aix Marseille II/França, Doutor em Informática - UM2/França.
marcello.bax@gmail.com

Como citar este artigo:

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