

Transparência para Humanos e Máquinas: Um *framework* para Publicar Dados Abertos Interconectados Semanticamente Descritos

Julio Cesar dos Reis¹, Mario Ferreira de Brito¹

¹Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Av. Albert Einstein, 1251, Cidade Universitária, CEP 13083-852, Campinas, SP – Brasil

jreis@ic.unicamp.br, profmariobrito@gmail.com

Abstract. *The Semantic Web allows data semantics to be explicitly described for humans and machines. Data transparency requires the structured publication of data for other systems. Open and interconnected data made available on the Web without copyright restrictions is key to achieving transparency mechanisms. However, often data to be published by organizations is found in many isolated systems that do not interact with each other. In this article, we propose a framework to enable the publication of open interconnected data, starting from several data sources for transparency. The study context is in a public university in Brazil where data of different nature need to become available to users of different profiles.*

Resumo. *A Web Semântica permite que a semântica dos dados seja descrita explicitamente para humanos e máquinas. Transparência de dados requer a publicação estruturada dos dados para outros sistemas. Dados abertos interconectados que são disponibilizados na Web sem restrições de “copyright” são a chave para atingirmos mecanismos de transparência. Mas muitas vezes os dados a serem publicados pela organização se encontram em diversos sistemas isolados. Neste artigo, propomos um framework para possibilitar a publicação de dados interconectados abertos, partindo de diversas fontes de dados visando a transparência. O contexto de estudo é em uma universidade pública onde dados de natureza distinta necessitam ser disponibilizados aos usuários de diferentes perfis.*

1. Introdução

A Web tem evoluído de documentos em hipertexto conectados para conceitos e dados interconectados que possibilitam integração de dados heterogêneos entre múltiplos sistemas de informação. Como evolução natural da Web, a Web Semântica investiga métodos visando a modelagem e descrição de dados com definições semânticas associadas [Berners-Lee 2001]. Nesse contexto, o conceito de *Linked Data* define uma série de princípios e técnicas para a publicação e o acesso de dados estruturados em repositórios Web [Wood 2014]. O principal objetivo é possibilitar a interligação de diversas fontes de dados e alavancar o uso e análise de dados.

Transparência refere-se a noção de tornar público informações de interesse da sociedade tornando visível informação a um ou mais interessados [Engiel 2015]. A Web se

tornou um ambiente propício para a publicação e consumo de dados abertos para promover transparência. Atualmente há diversos modelos de dados e princípios para a criação e disponibilização de dados abertos [Dadzie 2011].

Muitas iniciativas visando a transparência de dados públicos e governamentais têm adotado a Web como plataforma de compartilhamento. Essas iniciativas têm incentivado a disponibilização de dados abertos com o propósito de tornar o acesso aos dados sem restrições e transparentes [Janev 2012]. Esse movimento é principalmente observado com dados governamentais ou dados de pesquisa para permitir que a população ou outros usuários interessados tenham acesso transparente aos fatos das instituições e acompanhe a mesma conforme seus interesses.

No entanto, usualmente diferentes sistemas computacionais definem seus próprios modelo de dados heterogêneos em diversificados formatos, criando sérias dificuldades para a interoperabilidade e entendimento do significado dos dados por humanos e máquinas [Bizer 2009]. Esta pesquisa aborda a problemática de publicar dados abertos e semanticamente descritos com tecnologias da Web Semântica a partir de diversas fontes de dados. Os desafios envolvem o entendimento de modelo de dados de diferentes sistemas que codificam dados completamente distintos, de modo que seja possível entender o significado dos mesmos. O contexto prático de dados em universidade pública se caracteriza com sistemas computacionais dessa natureza.

Embora a literatura tenha abordado infraestruturas e aplicações para a publicação de dados interconectados abertos [Wood 2014], é necessário estudar técnicas que permitam estabelecer um processo desde a origem até a publicação, considerando a atualização dos dados periodicamente na Web. Isso envolve a criação de modelos para a descrição semântica e métodos para a geração e transformação automática de dados de um modelo para outro. Adicionalmente, os desafios envolvem a criação de endereços de identificação únicos e a criação automática de mapeamentos explícitos com conjunto de dados existentes. Em particular, mecanismos de atualização dos dados publicados precisam ser estudados para permitir consistência ao longo do tempo.

Este artigo investiga conceitos e técnicas para a publicação e atualização de dados interconectados abertos com semântica interpretável pela máquina. Propomos um *framework* computacional, nomeado de “*LODCamp*”, que permite gerar e atualizar dados interconectados para disponibilizar repositórios de dados estruturados e semanticamente descritos na Web visando a transparência de dados na universidade. O *framework* explora tecnologias da Web Semântica como RDF(s) e linguagens de descrição de ontologias OWL.

A metodologia de pesquisa envolve diversos passos iniciando-se pela revisão detalhada da literatura com o objetivo de entender os estudos e técnicas existentes. Isso também envolve uma análise minuciosa de infraestrutura e aplicações no contexto desta proposta. Em seguida, estudamos técnicas e modelos para a publicação e integração de dados abertos na Web Semântica.

Com base nesse contexto, e como nossa principal contribuição definimos um *framework* original para a publicação de dados abertos interconectados. Exploramos dados de uma universidade pública, a Universidade Estadual de Campinas (UNICAMP) em Campinas, São Paulo, para entender os desafios e prospectar soluções. Transparência é

relevante para a universidade e órgãos governamentais, e poucos trabalhos exploram tecnologias da Web Semântica para publicação de dados abertos nesse contexto, sendo esta a principal motivação para a condução dessa investigação.

O restante deste artigo está organizado da seguinte maneira: A seção 2 apresenta os conceitos fundamentais e trabalhos relacionados. A seção 3 descreve nosso *framework*, enquanto a seção 4 discute os resultados obtidos. A seção 5 finaliza o artigo com as considerações finais.

2. Trabalhos Relacionados

O formato heterogêneo de dados publicados na Web dificulta a interpretação desses dados e a ligação com outros conjuntos de dados disponíveis. Como meio de mitigar esse problema, propostas têm definido modelos de dados e técnicas que permitam a publicação e acesso a dados interconectados um com outros [Bizer 2009].

O modelo RDF baseia-se em identificar elemento de dados por meio de URI's, como identificadores únicos para os dados na Web, e descreve a relação entre dados via outras URI's. Essas são as chamadas triplas RDF (*Resource Description Framework*), em que se tem-se um sujeito, e um objeto que são ligados por um predicado. Com base no modelo de dados RDF, a proposta de “*Linked Data*” (dados ligados) [Bizer 2009] visa definir um conjunto de técnicas para a publicação de dados na Web, combinando e interligando diversos formatos e interfaces padrões com diversas fontes de diferentes localidades.

Berners-Lee *et al.* [Berners-Lee 2001] definiu um conjunto de princípios para a publicação de dados abertos na Web. Os princípios envolvem o uso de URIs para acesso global aos dados, uso de formatos não proprietários para a descrição dos dados, e sublinha a relevância da conexão entre elementos de dados distintos. De fato, o principal benefício é haver ligações explícitas e semanticamente definidas entre elementos de dados de maneira global, abertamente disponíveis na Web. Dadzie & Rowe [Dadzie 2011] discutiram os princípios de *Linked Data* na Web em grande escala e apontaram o grande interesse despertado pelas grandes empresas e entidades governamentais.

A semântica formalmente descrita e interpretável por máquina é obtida através do uso de artefatos como ontologias. Essas codificam o conhecimento interpretável por computador [Breitman 2007], tornando explícito para máquinas a semântica dos dados, que, por sua vez, facilita sua exploração nessa direção.

Janev *et al.* [Janev 2012] discutiram a evolução de dados abertos governamentais ao público em geral. Esse trabalho fez a extração de dados de mais de 100 servidores da Sérvia (antiga Iugoslávia) usando como ferramentas modelos RDFs e linguagem SPARQL¹, uma linguagem da Web Semântica que permite efetuar consultas em bases RDF. Nesse projeto, obteve-se a extração de preços, contas governamentais, usos de informações e tecnologias, assim também como atividades no geral realizadas pelo governo. Os autores concluíram que o modelo RDF trouxe benefícios para a publicações de dados nesse estudo de caso, permitindo uma grande quantidade de informação ser organizada e publicada de maneira interoperável na Web.

¹<https://www.w3.org/TR/rdf-sparql-query/>

Batista *et al.* [Batista 2013] argumentaram sobre a expansão da web semântica e usos das tecnologias RDF e OWL com *Linked Data* para divulgação dos dados. Apresentaram um estudo de caso sobre a publicação de dados abertos sobre os artigos do Simpósio Brasileiro de Banco de Dados. Definiram ontologias para a publicação do conjunto de dados permitiram a consulta aos mesmos via a linguagem SPARQL.

Bohm *et al.* [Böhm 2012] desenvolveram um sistema para coleta e disponibilização de dados abertos governamentais em larga escala. O sistema com interface Web permite usuários acessarem os dados e criar questões sobre políticos e fundos governamentais. As respostas são procuradas com base nos dados coletados e disponibilizados via formato RDF. Ainda nesse contexto, Moraes *et al.* apresentaram uma arquitetura para publicação e análise visual de dados interconectados usando padrão RDF visando aprimorar a gestão de dados na saúde pública.

Keßler & Kauppinen [Kessler 2015] indicaram que a publicação de dados abertos semanticamente descritos é essencial para a transparência na administração de Universidades. Eles definiram uma infraestrutura para a publicação de dados abertos da universidade de Münster, na Alemanha. Argumentaram que tal infraestrutura desenvolve um papel central para a visibilidade e acessibilidade dos dados produzidos e coletados pela universidade.

Fagundes & Sampaio [Fagundes 2017] discutiram a gestão pública administrativa e as dificuldades do cidadão no acesso as informações, tendo como base o portal da participação social “Participa.br”. Os autores indicaram que a sociedade brasileira está mudando de uma maneira a exigir dos órgãos governamentais, mais transparência em termos de gestão pública, possibilitando desse modo o acesso às informações pelo cidadão.

Silva *et al.* [Silva 2017] avaliaram o acesso às informações governamentais, em especial dos institutos federais da região centro oeste do Brasil. Os autores indicaram as dificuldades que brasileiros enfrentam no manuseio das informações disponibilizadas para visualização de dados governamentais. Concluem o trabalho mostrando através da pesquisa, que muitas unidades governamentais do centro oeste do Brasil não estão preparadas de acordo com a lei de transparência governamental reconhecida e aceita no Brasil.

Nossa análise da literatura indicou poucos trabalhos que visam explorar dados ligados descritos e publicados em RDF para integração de informação em universidades públicas. Nesta pesquisa, contribuímos com técnicas para a descrição, geração e atualização de dados heterogêneos descritos semanticamente via um *framework* de *software* para consumo de dados de diferentes fontes e geração automática de dados interconectados abertos. Acreditamos que nossa proposta inova no fato de permitir o usuário acessar informação organizada a partir de diferentes fontes. Exploramos dados da UNICAMP como estudo de caso e experimentação nesse trabalho.

3. Framework LODCamp

A figura 1 apresenta nossa proposta do *framework* onde os principais componentes envolvidos no sistema são destacados. Esse *framework* consiste de uma integração entre cinco módulos principais que são; A - as bases de dados da universidade (entrada principal de dados); B - módulo de acesso à vocabulários externos de domínios universitários; C - composição dos elementos de solução para a integração de dados (principais componen-

tes da solução); D - serviços e aplicações para consumo dos dados RDF; e E - base RDF para disponibilizar dados abertos (sistema Virtuoso para armazenamento de RDF).

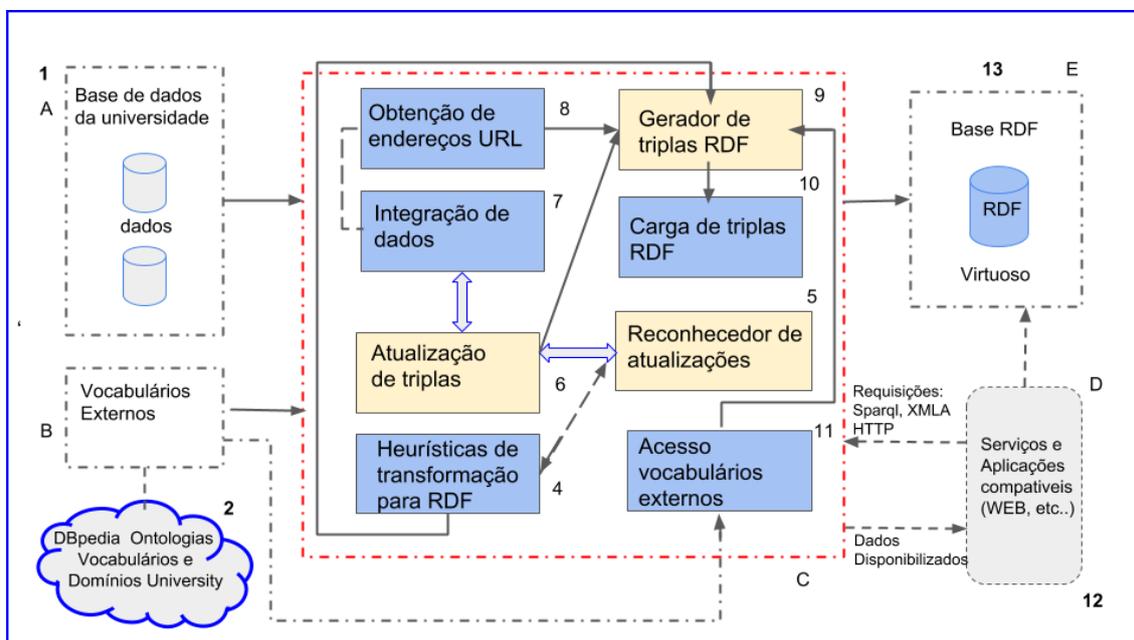


Figura 1. Componentes do *Framework LODCamp*

Inicialmente os dados da universidade se encontram disponíveis em diversas *Bases de dados da universidade* e outros tipos de arquivos que organizam dados. Esta é a principal fonte de dados para o *framework* e o início do processo de divulgação e transparência de dados. Consideramos como entrada arquivos que estão em diversos formatos tais como: .csc, .xml, .sql, .odf, etc., e devem ser tratados para ser disponibilizados aos usuários e sistemas finais.

Igualmente como entrada para o *Framework* estão os *Vocabulários externos* (B na Figura 1). O *LODCamp* deve acessar vocabulários estruturados que definem a semântica de dados do domínio. Para esse fim, consideramos ontologias sobre o domínio universitários e outras bases de conhecimento relevantes e ontologias que definem pessoas, e.g., *FOAF*². Por exemplo, a *DBpedia*³, como uma base de conhecimento que descreve artigos da Wikipédia, pode ser uma fonte para interligarmos dados da universidade com conjunto de dados externos, enriquecendo a divulgação. O *framework* define uma interface de software para permitir acesso à dados ligados disponíveis na Web de forma geral através de endereços URI e mecanismos de consulta de dados via SPARQL. A proposta considera o uso de vocabulários padrões já extensamente explorados e visa definir novas ontologias (usaremos a língua Portuguesa) para suprir as necessidades de publicação de dados na UNICAMP.

Os componentes do *LODCamp* (C na Figura 1) compõem os elementos para o tratamento dos dados. Esse módulo é responsável pela transformação dos dados de entrada em RDF. Inicialmente, O componente *reconhecedor de atualizações* (5) é responsável por

²<http://www.foaf-project.org/>

³<http://pt.dbpedia.org/>

monitorar a entrada de dados e reconhecer atualizações (novos dados ou dados modificados) advindos das fontes de dados da universidade. Uma vez reconhecida a atualização através do reconhecedor de atualizações, inicia-se o processo de atualização de triplas *Atualização de triplas* (6). Esse componente interage com o de Integração de Dados (7) e o Gerador de triplas RDF (9).

Uma grande problemática é efetuar e tratar a decisão entre triplas que representam elementos de dados existentes, e que devem ser modificadas, devido à alguma modificação no dado original, ou a criação de uma tripla totalmente nova, cujo dado não existia em formato RDF antes. A *Atualização de triplas* (6) é responsável por esta tarefa e a Integração de Dados (7) efetua a relação da tripla nova (atualizada) com outros elementos de dados. Por exemplo, um dado novo pode ser ligado a algum fato já existente na base RDF (interna ou externa).

O Gerador de triplas RDF (9) é responsável pela criação de fatos na base de conhecimento em RDF. Para esse fim, ele depende de três componentes chaves: *heurística de transformação para RDF* (4), Acesso à vocabulários externos (11) e obtenção de endereços URL (8). O componente de heurística é responsável por definir diretrizes de como as triplas serão definidas a partir de diversos formatos de dados de entrada. Nesse momento, é necessário o uso de vocabulários externos com as ontologias que definem as classes para representar a semântica dos dados.

A geração de uma nova tripla exige a definição de um endereço único para o item de dado. Para esse fim, o componente obtenção de endereços URI (8) é responsável por criar URI na geração de triplas. O sistema então verifica se houve acréscimo de arquivo ou se simplesmente foi substituído um arquivo existente. Caso seja um arquivo existente, o mesmo já possui um endereço URL e então este endereço é disponibilizado para o gerador de triplas. Caso contrário, caso seja um dado inteiramente novo, o sistema gera um novo endereço URI, disponibilizando-o e enviando uma informação ao gerador de triplas.

O componente de carga de triplas (10) efetua a transformação de fato dos dados em seu formato original para os fatos em RDF conforme a tripla gerada em (9) que são armazenados na base de triplas virtuoso⁴. As triplas conectam-se umas com as outras formando um grafo, e definindo os endereços URI de modo que possam ser acessados pelos sistemas finais de busca.

Os *serviços e aplicações* (D na Figura 1) são os sistemas de acesso as informações do módulo de integração de dados *LODCamp*. Nesse módulo, temos os navegadores Web e diversas requisições tais como HTTP e interfaces como o SPARQL para a requisição de dados no sistema.

Elaboramos um exemplo para ilustrar as etapas principais do *framework*. A Figura 2 apresenta um exemplo de dados de entrada na base inicial em uma planilha. Os dados brutos não indicam o significado dos itens e são necessários vocabulários externos para associar com os dados e possibilitar a descrição explícita da semântica.

O *framework* requer a criação de uma “URI”⁵ de base, nos quais os dados são

⁴<https://virtuoso.openlinksw.com/rdf/>

⁵Como exemplo ilustrativo, considere uma URI de base: `;;http://universidade/resources;ç;ç`. Usamos a sintaxe `@prefix uni: ;;http://universidade/resources;ç;ç` para se referir a esta URI.

	A	D	E	F	G	H	I	J	K
1	ALUNOS ATENDIDOS	2009	2010	2011	2012	2013	2014	2015	2016
2	Alunos atendidos com Bolsa Alimentação e Transporte	854	933	1.396	1.093	955	1.074	1.147	1.057
3	Alunos atendidos com Bolsa Alimentação e Transporte	30	30	30	30	38	32	30	30
4	Alunos atendidos com Bolsa Alimentação e Transporte	-	-	109	229	271	266	364	275
5	Alunos atendidos com Bolsa Auxílio Emergência	251	370	449	403	259	255	219	237
6	Alunos atendidos com Bolsa Auxílio Estudo Formação	-	-	-	-	-	30	78	69
7	Alunos atendidos com Bolsa Auxílio Instalação	-	-	-	-	111	193	112	200
8	Alunos atendidos com Bolsa Auxílio Intercâmbio	40	40	20	78	43	53	52	45
9	Alunos atendidos com Bolsa Auxílio Moradia	-	177	272	354	416	474	603	583
11	Alunos atendidos com Bolsa Auxílio Social	1.048	1.046	1.108	1.234	1.232	1.466	1.655	1.722

Figura 2. Dados iniciais de entrada (estruturados em uma planilha)

concatenados. Ela será usada como meio de identificar unicamente os elementos de dados que serão transformados em RDF.

O *framework* transforma a planilha de entrada em triplas RDF. Para isso exploramos a linguagem “Turtle”⁶ como meio de expressar as triplas RDF⁷. A Figura 3 apresenta duas triplas de exemplo que são geradas pelo *framework* com base na planilha de entrada. A Figura 3 (a) descreve que o sujeito “uni:bolsa_auxilio_moradia” é do tipo (uso da propriedade *rdfs:type*) “onto:bolsa”. Nesse caso, essa URI tem como prefixo “onto” que define uma ontologia do contexto universitário. Logo, “onto:bolsa” é uma URI da classe que representa o conceito de bolsa na universidade. Nesse contexto, a aplicação pode interpretar que “uni:bolsa_auxilio_moradia” é uma bolsa de estudo, e não uma bolsa, como utensílio de guardar objetos.

A Figura 3 (b) descreve o fato em uma tripla RDF em que “uni:bolsa_auxilio_moradia” está ligado via a propriedade “onto:ano_2010” ao valor 177. Nesse caso, estamos usando o @prefix *xsd*=”http://www.w3.org/2001/XMLSchema» para expressar que 177 é um valor decimal via o *schema xmls* já existente. A proposta permite reuso de vocabulários criados e disponíveis na Web, como esse exemplo de uso do vocabulário que expressa tipos de dados em XML.

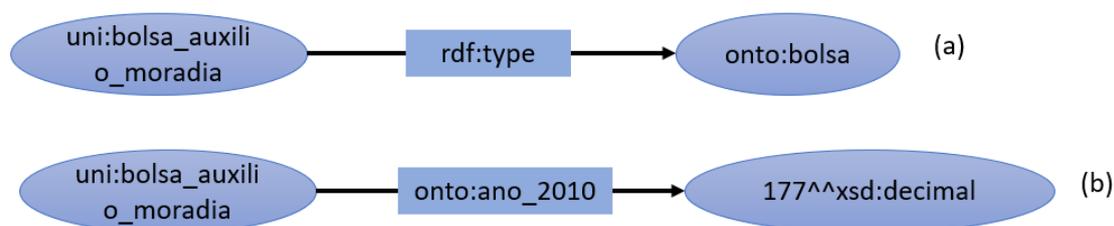


Figura 3. Exemplo de triplas RDF geradas com o uso da linguagem Turtle

Essas triplas são armazenadas na base de triplas RDF (virtuoso) para ser acessado por outras aplicações por meio de consultas e propostas de visualização de dados. Nosso exemplo objetivou apresentar uma visão geral das entradas e saídas do *framework* de modo a permitir acesso semântico aos dados da universidade. Elucidamos os passos de modo a comunicar o mecanismo de publicação de dados abertos através de princípios *Linked data* e uso de modelo de dados RDF.

⁶<https://www.w3.org/TR/turtle/>

⁷“Turtle” possui um formato simples em RDF para leitura final das pessoas

4. Discussão

Linked data é uma proposta que permite tornar explícito a semântica de dados publicados na Web e conseqüentemente alavanca a integração e recuperação de informação. Essa proposta desenvolve um papel chave em ser uma via para tornar mais acessível dados para pessoas e máquinas. É extremamente relevante tornar possível aos governos e administrações públicas a divulgação de dados públicos de uma maneira inovadora, de modo que usuários consigam acessá-los e usá-los. No entanto, argumentamos que simplesmente disponibilizar documentos na Web não significa que as instituições estão possibilitando transparência de dados. É preciso tratar os dados e disponibilizá-los de modo que usuários humanos e máquinas possam interpretá-los de uma maneira adequada.

Nesta pesquisa propomos uma infraestrutura como uma solução técnica para divulgar dados abertos semanticamente através de princípios de *Linked Data* e modelo de dados em RDF. Nossa inovação consistiu sobre a definição de um conjunto de componentes interligados no *framework LODCamp* que permitem a partir de dados de diferentes formatos e natureza serem publicados em uma base de dados RDF unificada. Em termos de transparência para humanos, entendemos que dados descritos em RDF não sejam apresentados diretamente para os usuários. Os dados brutos em RDF são adequados para a integração de dados entre sistemas, e o acesso dos usuários requer um mecanismo de visualização visando promover a transparência e entendimento facilitado dos dados pelo cidadão.

O *LODCamp* define componentes que atuam em todo o processo de publicação de dados, desde a chegada dos mesmos na base de dados da universidade, até sua saída para divulgação em formato RDF. O *LODCamp* atualiza os dados, efetuando a transformação e carregamento das triplas, entendendo as requisições, localizando os dados e disponibilizando-os aos sistemas finais para consultas. Trabalhos futuros envolvem investigar meios de permitir a visualização simplificada de triplas RDF.

Neste trabalho, analisamos o fluxo de divulgação de dados de uma maneira geral, desde a entrada de dados como fontes de dados da universidade (que é nosso caso) até a publicação final dos mesmos. Nosso *framework* pode ser base para construção de sistemas de *software* em outros tipos de órgãos governamentais para divulgar e atualizar ao longo do tempo a publicação de dados abertos na Web semanticamente descritos. Isso pode facilitar o acompanhamento pela população da administração pública governamental.

Como pontos que ainda necessitam ser aperfeiçoados nessa pesquisa são os mecanismos de monitoramento de atualização e as heurísticas de geração de RDF. Para esse fim, conduziremos observações em diversos dados na UNICAMP para entender os diferentes tipos e necessidades para atualização dos dados. Essa observação também beneficiará meios de formalizarmos diretrizes para as heurísticas.

Trabalhos futuros envolvem igualmente a definição de consultas e filtros adequados para acessos aos dados ligados integrados. Adicionalmente, prevemos desenvolver inteiramente o *framework* e conduzir estudos com dados reais da UNICAMP para avaliar a efetividade dos mecanismos de publicação e atualização de dados ao longo do tempo.

Acreditamos que explorar tecnologias já padronizadas da Web Semântica como RDF(S), SPARQL, Turtle, e plataformas existentes como o Virtuoso deva beneficiar o processo de publicação de dados abertos. Essas tecnologias desenvolvem um papel fun-

damental para a construção do *LODCamp*, fator esse relevante no acesso e divulgação dos dados abertos.

5. Conclusão

Transparência de dados em um contexto de diversidade em termos sociais e técnicos, como em uma universidade, demanda novas abordagens e suporte computacional para a publicação de dados que façam sentido para humanos e máquinas. Esta pesquisa contribuiu no desenvolvimento de técnicas para a geração de dados públicos e definidos semanticamente para pessoas e sistemas de *software* visando a transparência. Estudamos a construção de um *framework* que permita publicar dados interconectados abertos a partir de diversas fontes de dados heterogêneas advindos de diferentes setores da universidade. Descobrimos os elementos-chaves que devem compor uma solução para publicação de dados semânticos na universidade. Ilustramos o uso do *framework* abordando os desafios de entender e definir modelos comuns para a descrição semântica de dados de diferentes fontes de informação de uma universidade pública, a UNICAMP, através do *framework*. Trabalhos futuros envolvem o desenvolvimento de ferramentas de *software* que implementam os conceitos definidos nesta investigação.

Agradecimentos

Esta pesquisa tem apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Processo #2017/02325-5).

Referências

- Batista, Mateus Gondim Romão e Lóscio, B. F. (2013). Usando linked data para a publicação de dados abertos sobre o sbbd. In *XXVIII Simpósio Brasileiro de Banco de Dados, Recife, Pernambuco, Brasil*, pages 10:1–10–6.
- Berners-Lee, Tim e Hendler, J. e. L. O. (2001). The semantic web. *Scientific American*, 284(5):34–43.
- Bizer, C. e Heath, T. e. B.-L. T. (2009). Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22.
- Breitman, Karin Koogan e Casanova, A. M. e. T. W. (2007). *Semantic Web: Concepts, Technologies and Applications*. Springer Verlag, London.
- Böhm, Christoph e Freitag, M. e. S. M. e. o. (2012). Govwild: Integrating open government data for transparency. Anais da Conferência WWW.
- Dadzie, Aba-Sah e Rowe, M. (2011). Approaches to visualising linked data: A survey. *Semantic Web*, 2(2):89–124.
- Engiel, Priscila, P. J. P. R. e. L. J. C. S. d. P. (2015). Representando o conceito de transparência através de um léxico ampliado da linguagem. WTranS - Workshop de Transparência em Sistemas.
- Fagundes, Vladimir e Sampaio, J. d. O. (2017). Análise da plataforma de participação social: Participação do cidadão e organizações na elaboração de políticas públicas. XXXVII Congresso da Sociedade Brasileira de Computação (5^oWTranS).

- Janev, Valentina e Milošević, U. e. o. (2012). Linked open data infrastructure for public sector information: Example from serbia. Anais da Conferência Internacional I-SEMANTICS.
- Kessler, Carsten e Kauppinen, T. (2015). Linked open data university of münster – infrastructure and applications. In *The Semantic Web: ESWC 2012 Satellite Events*, pages 447–451. Springer, Berlin, Heidelberg, Berlin, Heidelberg.
- Silva, Camila L. e Felberg, J. A. e. C. J. e. L. A. C. (2017). Um panorama sobre a acessibilidade para transparência em sites de institutos federais do centro-oeste do brasil. XXXVII Congresso da Sociedade Brasileira de Computação (5ºWTranS - Workshop de Transparência em Sistemas).
- Wood, David e Zaidman, M. e. R. L. (2014). *Linked Data: Structured Data on the Web*. Manning Publications CO.