

Revista Eletrônica de Sistemas de Informação

ISSN 1677-3071

V. 10, n. 1

2011 - Edição temática sobre governo eletrônico

doi:10.5329/RESI.2011.1001

Sumário

Editorial

SOBRE ESTA EDIÇÃO

Rodrigo Sandoval Almazán, Ernani Marques da Silva, Alexandre Reis Graeml

RESI NO QUALIS (2)

Alexandre Reis Graeml

E-gov mundo a fora

COAXING AN INFORMATION SOCIETY IN THE DOMINICAN REPUBLIC: THE RISE AND STEEP FALL OF A TECHNOLOGY PARK'S UNIVERSITY RESEARCH CENTER

Julio Angel Ortiz

ALGUNAS NOTAS SOBRE PARTICIPACIÓN ELECTRÓNICA EN ESPAÑA. DOS EXPERIENCIAS REALES EN EL AÑO 2010: CADRETE (ZARAGOZA) Y BARCELONA

José María Moreno Jiménez, Manoela Velázquez Arguedas

E-gov no Brasil

PORTAIS DE SERVIÇOS PÚBLICOS E DE INFORMAÇÃO AO CIDADÃO NO BRASIL: UMA DESCRIÇÃO DO PERFIL DO VISITANTE

Maria Alexandra Viegas Cortez da Cunha, José Roberto Frega, Iomara Scandelari Lemos

COMPRAS ELETRÔNICAS GOVERNAMENTAIS: UMA AVALIAÇÃO DOS SITES DE E-PROCUREMENT DOS GOVERNOS ESTADUAIS BRASILEIROS

Tomaz Rodrigo Alves, Cesar Alexandre Souza

INICIATIVAS DE GOVERNO ELETRÔNICO: ANÁLISE DAS RELAÇÕES ENTRE NÍVEL DE GOVERNO E CARACTERÍSTICAS DOS PROJETOS EM CASOS DE SUCESSO

Edmir Parada Vasques Prado, Neilson Carlos Leite Ramalho, Cesar Alexandre de Souza, Maria Alexandra Viegas Cortez da Cunha, Nicolau Reinhard

Foco na tecnologia

AUMENTANDO A TRANSPARÊNCIA DO GOVERNO POR MEIO DA TRANSFORMAÇÃO DE DADOS GOVERNAMENTAIS ABERTOS EM DADOS LIGADOS

Lucas de Ramos Araújo, Jairo Francisco de Souza

DETECÇÃO DE CARTÉIS EM LICITAÇÕES PÚBLICAS COM AGENTES DE MINERAÇÃO DE DADOS

Carlos Vinícius Sarmiento Silva, Célia Ghedini Ralha



Este trabalho está licenciado sob uma [Licença Creative Commons Attribution 3.0](http://creativecommons.org/licenses/by/3.0/).

ISSN: 1677-3071

Esta revista é (e sempre foi) eletrônica para ajudar a proteger o meio ambiente, mas, caso deseje imprimir esse artigo, saiba que ele foi editorado com uma fonte mais ecológica, a *Eco Sans*, que gasta menos tinta.

AUMENTANDO A TRANSPARÊNCIA DO GOVERNO POR MEIO DA TRANSFORMAÇÃO DE DADOS GOVERNAMENTAIS ABERTOS EM DADOS LIGADOS

INCREASING GOVERNMENT TRANSPARENCY BY TRANSFORMING OPEN GOVERNMENT DATA INTO LINKED DATA

(paper submitted in April 2011)

Lucas de Ramos Araújo

Departamento de Ciência da Computação – Universidade Federal de Juiz de Fora (UFJF)
lucas.araujo@ice.ufjf.br

Jairo Francisco de Souza

Departamento de Ciência da Computação – Universidade Federal de Juiz de Fora (UFJF)
jairo.souza@ufjf.edu.br

ABSTRACT

Since its inception, the Web has undergone constant evolution in order to improve itself as a mean of global communication and information sharing. Open Government Data are increasingly being published on the Web, contributing to the transparency and the reusability of public data. At the same time, the use of Linked Data has been increasing in recent years, enabling the development of better and smarter applications. In this context, this paper proposes the publication of Open Government Data using the Linked Data practices, by creating a data set of Brazilian politicians with information collected from different sources.

Key-words: e-government; linked data; semantic web.

RESUMO

Desde o seu surgimento, a Web vem sofrendo constantes evoluções a fim de se aprimorar cada vez mais como meio de comunicação global e compartilhamento de informações. Dados Governamentais Abertos estão sendo cada vez mais publicados na Web, contribuindo para a transparência e a sua reutilização. Ao mesmo tempo, a prática de disponibilizar dados em formato RDF com significados explicitamente definidos e com ligações para outros dados externos, conhecida como dados ligados, vem crescendo muito nos últimos anos, permitindo o desenvolvimento de aplicações melhores e mais inteligentes. Neste contexto, este trabalho tem como proposta apresentar o projeto "Ligado nos Políticos", que utilizou práticas de dados ligados para criar um conjunto de dados abertos de políticos brasileiros com informações coletadas de diferentes fontes.

Palavras-chave: governo eletrônico; dados ligados; web semântica.

1 INTRODUÇÃO

As Tecnologias da Informação e de Comunicação (TICs) promoveram uma revolução nos meios de informação, construindo uma nova relação entre governo e cidadãos. Esta nova relação deu origem ao chamado Governo Eletrônico, que possibilita uma administração pública mais acessível, eficiente, democrática e transparente. Dentro deste contexto, o conceito de Dados Governamentais Abertos surge para ampliar esta relação, promovendo a disponibilização das informações governamentais em formatos abertos e acessíveis a fim de possibilitar a reutilização e a interligação com informações de outras fontes, gerando novos significados (W3C Escritório Brasil, 2010).

Atualmente, muitos dados governamentais estão disponíveis na Web, mas estas informações, na maioria das vezes, são oferecidas sem a utilização de padrões, em formatos proprietários ou apenas para a visualização, dificultando a reutilização. Para bem aproveitar o potencial representado pelo acervo de informações do governo, essas informações precisam ser disponibilizadas em formato padronizado, aberto e acessível (AGUNE *et al.*, 2009).

Existem diversas maneiras de publicar Dados Governamentais Abertos, mas segundo Berners-Lee (2009) os objetivos esperados ao publicar dados governamentais são mais bem alcançados usando-se dados ligados. Dentro do contexto da Web Semântica, o termo dados ligados (*linked data*) é utilizado para descrever um conjunto de práticas para publicar, compartilhar e conectar dados estruturados na Web de forma a aumentar o seu valor e utilidade (BIZER *et al.*, 2009).

Há um movimento cada vez maior de governos, organizações e pessoas publicando Dados Governamentais Abertos. Ao mesmo tempo, a utilização de dados ligados vem crescendo muito nos últimos anos, sendo fortemente apoiada pelo W3C (*World Wide Web Consortium*) e por Tim-Berners Lee, considerado o inventor da Web (W3C, 2011).

Porém, vários desafios ainda devem ser superados para que a Web seja utilizada como um grande banco de dados global. Segundo o relatório United Nations E-Government Survey 2010 (UNITED NATIONS, 2010), que apresenta a situação mundial no setor de Governo Eletrônico, o Brasil ocupa a posição de número 61, acumulando uma perda de 16 posições desde 2008. Diversos fatores são responsáveis pelo declínio brasileiro, tais como a insuficiência de serviços *online* e a deficiente infraestrutura de telecomunicações. O relatório destaca ainda iniciativas brasileiras de dados abertos que devem ser seguidas.

Dentro do contexto apresentado, pode-se perceber a importância da publicação de Dados Governamentais Abertos e a relevância das práticas de dados ligados na Web atual. Este artigo tem como objetivo apresentar o trabalho realizado para publicação de dados ligados de políticos brasileiros na Web, através da criação de um *data set* com informações coletadas de diferentes fontes, contribuindo assim com a nova Web de

dados. O projeto implementado tem como objetivo fornecer dados úteis, abertos, padronizados, reutilizáveis e ligados a dados de outras fontes.

O restante deste trabalho está estruturado da seguinte maneira: a seção 2 discute sobre dados governamentais abertos e a seção 3 discute sobre a prática de dados ligados. Em seguida, a seção 4 apresenta uma descrição do trabalho realizado para criação do sistema Ligado nos Políticos, hoje o primeiro *dataset* de políticos brasileiros que disponibiliza suas informações usando dados ligados. Por fim, a seção 5 traz nossas conclusões juntamente com propostas de trabalhos futuros.

2 DADOS GOVERNAMENTAIS ABERTOS

Segundo o GI para e-Gov (2009), criar um Governo Eletrônico exige abertura, transparência, colaboração e conhecimento. Um governo transparente é mais do que a interação e a participação aberta; os dados do governo precisam ser partilhados, descobertos, acessíveis e manipuláveis por aqueles que os desejam para bem aproveitar as vantagens da Web e o acervo de informações das organizações.

A disponibilização de Dados Governamentais Abertos permite que os usuários possam facilmente encontrar, acessar, entender e utilizar os dados públicos segundo foco e interesses próprios, trazendo diversos benefícios como a reutilização, inclusão, transparência, responsabilidade, melhoria nas buscas, integração, participação, colaboração, crescimento econômico, inovação e eficiência (DINIZ, 2009).

O Open Government Working Group (OPENGOVDATA.ORG, 2007), elaborou os 8 Princípios dos dados governamentais abertos. Eles devem ser: (1) completos, totalmente disponíveis e sem limitações; (2) primários, coletados na fonte com o maior nível possível de granularidade, sem agregação ou modificação; (3) atuais, publicados tão rapidamente quanto necessário para preservar o seu valor; (4) acessíveis, disponibilizados para o maior número possível de usuários e finalidades; (5) processáveis por máquinas, razoavelmente estruturados para permitir processamento automatizado; (6) não-discriminatórios, disponíveis para todos sem necessidade de cadastro; (7) não-proprietários, disponibilizados em um formato sobre o qual nenhuma entidade tem controle exclusivo; (8) licenças livres, não sujeitos a nenhuma regulação de direitos autorais, patentes, propriedade intelectual ou segredo industrial.

As principais tecnologias e formatos utilizados para a publicação de dados governamentais abertos são: (1) arquivos CSV (*Comma Separated-Values*), que armazenam dados tabulares; (2) informações Atom e RSS (*Really Simple Syndication*), que agregam conteúdo baseado em XML, usadas para compartilhar novidades ou textos completos através dos denominados *feeds*; (3) interfaces REST, que associam um recurso a um URI usando HTTP, permitindo que um *site* possa ser enriquecido com aplicativos que expandam o valor de um recurso disponível; (4) tecnologias da web semântica, que oferecem um arcabouço comum onde os

dados podem ser compartilhados e reutilizados além dos limites de aplicativos, empreendimentos e comunidades (GI PARA E-GOV, 2009).

Há um movimento global de governos e autoridades locais disponibilizando seus dados na web. Projetos de dados governamentais abertos surgiram em vários países do mundo, como Estados Unidos, Reino Unido, Austrália, Nova Zelândia, Noruega, Holanda, Suécia, Espanha, Estônia, Áustria, Grécia, Canadá e Dinamarca, existindo também um número crescente de iniciativas locais de estados e cidades (SHERIDAN e TENNISON, 2010). Alguns governos criaram catálogos ou portais para tornar a localização e a utilização desses dados mais fácil para o público (BENNETT e HARVEY, 2009), como o portal data.gov e data.gov.uk. Além disso, pessoas e organizações vêm publicando dados governamentais por conta própria em vários formatos (BERNERS-LEE, 2009).

O Brasil tem uma boa oferta de dados em todas as esferas e poderes oferecidos pública e gratuitamente, mas existem poucas iniciativas do governo que se propõem a dar acesso à base integral estruturada e em linguagem aberta. O exemplo mais recente de iniciativa brasileira neste sentido é o projeto Governo Aberto SP, em fase de implantação (GOVERNO ABERTO, 2010). Enquanto o governo não libera mais dados em formato aberto, estão surgindo no Brasil iniciativas no sentido de extrair os dados de *sites* e portais governamentais, reorganizá-los, torná-los abertos e/ou conferir novo valor a eles através de diferentes aplicações, como o Congresso Aberto, o Parlamento Aberto, o Legisdados, entre outros (THACKER, 2011).

Dado o crescente interesse civil após exemplos bem sucedidos em outros países, espera-se que novas iniciativas sejam realizadas em esferas políticas brasileiras.

3 DADOS LIGADOS

Dados ligados referem-se a dados publicados na Web de modo que sejam legíveis por máquina, seus significados sejam explicitamente definidos, estejam ligados a outros conjuntos de dados e, por sua vez, possam ser ligados a partir de conjuntos de dados externos (BIZER *et al.*, 2009).

A idéia básica de dados ligados foi elaborada por Berners-Lee (2006), que definiu os quatro princípios que caracterizam os dados ligados e que devem ser aplicados para fazer a Web crescer semanticamente: (1) usar URIs para nomes de “coisas” (isto é, qualquer recurso); (2) usar URIs HTTP para que as pessoas possam procurar esses nomes; (3) fornecer informações úteis quando alguém acessar um URI, utilizando padrões como RDF (*Resource Description Framework*) e SPARQL (*SPARQL Protocol and RDF Query Language*); (4) incluir *links* para outros URIs para que as pessoas possam encontrar mais “coisas” (outros recursos). URIs, RDF e HTTP são as principais tecnologias utilizadas para disponibilizar dados ligados (BIZER *et al.*, 2008). Além dessas, outras tecnologias da Web Semântica são utilizadas para fornecer diferentes tipos de suporte, como a

linguagem SPARQL para a consulta de dados RDF, as linguagens RDFS (*Resource Description Framework Schema*) e OWL (*Web Ontology Language*) para a definição de vocabulários e a linguagem RDFa (*RDF-in-attributes*) para a publicação de dados com significado (HEATH & BIZER, 2011).

Os URIs identificam todos os itens de interesse na Web, chamados de recursos. Existem dois tipos de recursos: informacionais e não-informacionais. Todos os recursos que encontramos na Web tradicional, tais como documentos, imagens e outros arquivos de mídia, são recursos informacionais. Todos os “objetos do mundo real” que existem fora da Web são recursos não-informacionais, tais como pessoas, lugares, proteínas, conceitos científicos, entre outros (BIZER *et al.*, 2007).

RDF é um *framework* para representar informações na Web. A descrição de um recurso é representada como uma série de triplas com sujeito, predicado e objeto. O sujeito é o URI que identifica o recurso descrito. O objeto pode ser um valor literal, como uma *string*, número ou data, ou ainda o URI de um outro recurso que está relacionado ao sujeito. O predicado é um URI de algum vocabulário que indica o tipo de relação que existe entre o sujeito e o objeto (KLYNE e CARROLL, 2004).

Através do protocolo HTTP podemos acessar um URI na Web a fim de obter as informações sobre o recurso referenciado. Quando um URI que identifica um recurso informacional é desreferenciado, o servidor do URI gera uma representação do recurso e a envia para o cliente utilizando o código de resposta HTTP 200 OK. Recursos não-informacionais não podem ser desreferenciados diretamente. Uma das abordagens mais utilizadas é o redirecionamento HTTP 303, usado juntamente com a Negociação de Conteúdo. Clientes HTTP enviam cabeçalhos juntamente com cada solicitação para indicar que tipo de representação preferem. Servidores podem inspecionar os cabeçalhos e selecionar uma resposta adequada, em HTML ou RDF (BIZER *et al.*, 2007).

O exemplo mais visível da adoção e aplicação dos princípios de dados ligados é o projeto Linking Open Data, um esforço aberto e colaborativo apoiado pelo grupo W3C SWEO (*Semantic Web Education and Outreach Group*). O objetivo do projeto é identificar *data sets* existentes que estão disponíveis sob licenças abertas, convertê-los para RDF de acordo com os princípios de dados ligados, publicá-los na Web e interligá-los uns com os outros, formando uma nuvem de dados ligados (*LOD Cloud*) (BIZER *et al.*, 2008), conforme exemplo apresentado na Figura 1.

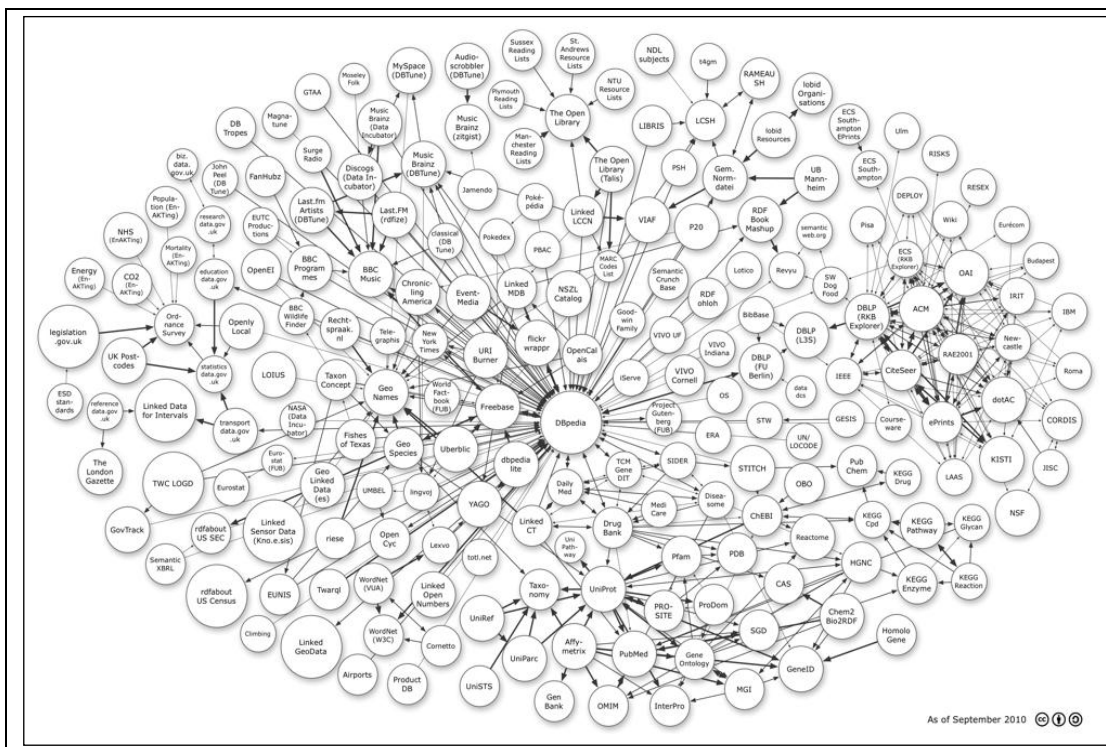


Figura 1. Diagrama da *LOD Cloud*

Fonte: Cyganiak e Jentzsch (2010)

Com um volume significativo de dados ligados sendo publicados na Web, inúmeras pesquisas e esforços estão sendo realizados para construir aplicações que explorem esta Web de dados: aplicações específicas através do *mashup* de dados de diferentes *data sets*, como o Reyvu e o DBPedia Mobile; motores de busca como o Falcons e indexadores como o Swoogle que navegam a Web de dados; e navegadores de dados ligados que permitem navegar entre fontes de dados seguindo links RDF, como o The Tabulator e o Disco Hyperdata Browser (BIZER *et al.*, 2009).

Ao publicar os dados ligados, URIs simples, pequenos, persistentes e estáveis devem ser escolhidos para os recursos a fim de que outros possam usá-los de forma segura. Além disso, deve ser considerada a infraestrutura técnica para torná-los desreferenciáveis. Geralmente são gerados três URIs relacionados a um único recurso não-informacional: um para o recurso; um para o recurso informacional adequado para navegadores HTML; e um para o recurso informacional adequado para navegadores RDF (BIZER *et al.*, 2007).

Para que aplicações clientes processem os dados mais facilmente, é considerada uma boa prática reutilizar termos de vocabulários bem conhecidos e amplamente utilizados sempre que possível. Novos termos só devem ser definidos se os termos necessários não puderem ser encontrados em vocabulários existentes. Nesse caso, deve ser publicado um arquivo usando as linguagens RDFS ou OWL (BERNERS-LEE, 2009). A fim de permitir que os clientes possam avaliar a qualidade dos dados

publicados e determinar se eles querem confiar neles, os dados devem ser acompanhados de vários tipos de metadados, como um URI que identifica o autor, a data de criação e o método de criação (BIZER *et al.*, 2008).

Os *links* RDF são a base dos dados ligados. Eles permitem que as aplicações cliente naveguem entre as fontes de dados e descubram dados adicionais. Para fazer parte da Web de Dados, fontes de dados devem definir *links* RDF para relacionar as entidades em outras fontes de dados (BIZER *et al.*, 2009). Links RDF podem ser definidos manualmente, mas é prática comum o uso de abordagens automatizadas ou semi-automatizada para gerar *links* RDF.

A fim de facilitar a descoberta dos dados, mecanismos adicionais podem ser utilizados, como adicionar o *data set* na Wiki ESW (W3C, 2010) para que outros desenvolvedores possam encontrá-lo mais facilmente, criar uma extensão *sitemap* para indicar onde o RDF está localizado e que meios alternativos são fornecidos para acessá-lo e definir links de páginas existentes para dados RDF. Além disso, é preciso testar se os dados estão publicados corretamente (HEATH *et al.*, 2008).

Na próxima seção, apresentaremos o trabalho realizado para publicar dados de políticos brasileiros utilizando a abordagem de dados ligados.

4 CRIAÇÃO DE UM *DATA SET* SOBRE POLÍTICOS BRASILEIROS

Para criação de uma base com dados de políticos brasileiros, foi necessário coletar informações de diversas fontes, agrupá-las e transformá-las em formato de dados ligados. A Figura 2 apresenta a arquitetura geral do projeto, mostrando os diferentes módulos e as ligações entre eles.

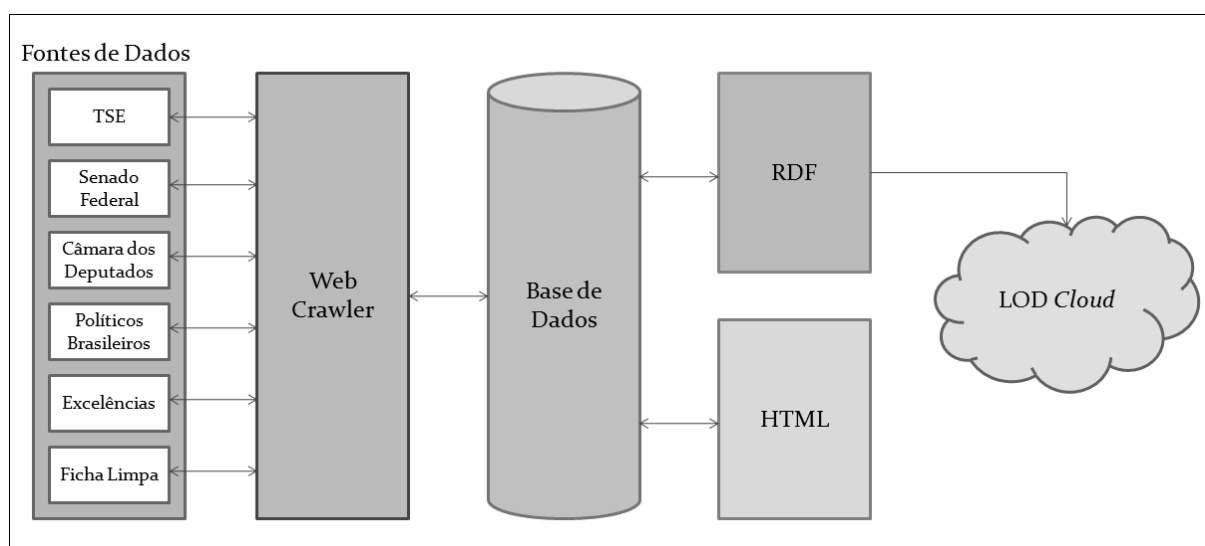


Figura 2. Arquitetura geral do projeto

Fonte: elaborada pelos autores.

Foi utilizado o nome “Ligado nos Políticos” para representar o *data set* e o domínio *http://ligadonospoliticos.com.br* para a publicação dos dados.

Como fontes de dados, foram utilizados o site do Tribunal Superior Eleitoral (TSE), o site do Senado Federal, o Portal da Câmara dos Deputados, o site da ONG (Organização Não Governamental) Políticos Brasileiros, o site Ficha Limpa e o projeto Excelências do *site* Transparência Brasil. Foram coletados dados pessoais, dados da eleição, divulgação de bens, dados parlamentares, lideranças, missões, mandatos, afastamentos, pronunciamentos, comissões, proposições e ocorrências.

Com exceção de alguns dados em formato CSV, a maioria dos sites disponibilizam as informações somente em formato HTML. Por isso, foi necessária a criação e utilização de *Web Crawlers* para extrair os dados de uma forma metódica e automatizada. A técnica para extrair dados *online* que não estão em formato aberto é denominada *screen scraping* ou raspagem de dados. Para cada fonte de dados, foi desenvolvido um *script* para coletar as informações relevantes e armazená-las em uma base de dados relacional.

As fontes que utilizamos não disponibilizam identificadores únicos para cada político. Para identificar os dados que tratavam da mesma entidade foi necessário utilizar os diferentes dados dos políticos, como nome, partido e data de nascimento nas cláusulas WHERE antes que uma nova entrada fosse inserida na base de dados, de forma a garantir a consistência e evitar a duplicação. A tarefa de melhor detectar se duas descrições de uma entidade referem-se à mesma entidade do mundo real é um problema de investigação conhecido como detecção de registro duplicado (ELMAGARMID *et al.*, 2007). Em alguns casos, foi identificada a falta de padrão na apresentação das informações nas diferentes fontes de dados. Por exemplo, o partido PC do B (Partido Comunista do Brasil) em alguns *sites* é apresentado como PCdoB; os Estados algumas vezes são representados através de siglas e outras através do nome. Portanto, esses dados tiveram que ser identificados e transformados para o formato de representação padrão escolhido por meio de cláusulas UPDATE.

Após obtermos todos os dados estruturados, a informação foi representada utilizando os princípios e práticas de dados ligados usando o modelo RDF. Foram escolhidos URIs HTTP simples e pequenas dentro do domínio para representar os recursos não-informacionais. Para garantir que cada URI fosse único, foi utilizada a chave-primária de cada político. Um exemplo de URI utilizado no projeto é *http://ligadonospoliticos.com.br/resource/1*, seguido de */html* ou */rdf*, dependendo da representação desejada.

Em seguida, foi realizada a escolha dos vocabulários que seriam utilizados para representar as propriedades dos recursos. Conceitos descritos em vocabulários conhecidos como FOAF¹, BIO², PERSON³, VCARD⁴,

1 <http://xmlns.com/foaf/0.1/>

DBPROP⁵, POL⁶, BEING⁷, TIMELINE⁸, DCTERMS⁹, MONEY¹⁰, GEOSPECIES¹¹, EVENT¹², SKOS¹³ e BIBLIO¹⁴ foram reutilizados sempre que possível. Os vocabulários FOAF, BIO, PERSON, GEOSPECIES, BEING e VCARD fornecem termos para descrição de dados pessoais e relacionamentos entre pessoas. O vocabulário POL fornece alguns termos para descrição de políticos. Os vocabulários EVENT e TIMELINE fornecem termos para descrição de eventos e relações temporais, respectivamente. Os vocabulários DCTERMS e SKOS fornecem termos para descrição de autorias e ligações entre documentos na web. O vocabulários MONEY fornece termo para descrição de moedas e transações financeiras. Por fim, o vocabulário DBPROP fornece termos descritos na base de dados da DBPedia, uma base dados de informações da Wikipedia em formato RDF. Para encontrar os termos foi utilizado o serviço *Talis Schemacache* e realizadas pesquisas a dados RDF de outros *data sets*. Não foram encontrados todos os termos necessários em outros vocabulários. Nesses casos, novos termos foram definidos no vocabulário representado por **polbr** sob o namespace <http://ligadonospoliticos.com.br/politicobr/>. A Figura 3 mostra o diagrama dos vocabulários e as respectivas propriedades utilizadas.

As informações são retiradas da base de dados e inseridas dinamicamente no modelo de acordo com o recurso solicitado. O sujeito ou recurso é representado pelo URI do político, os predicados ou propriedades pelos URIs dos termos preexistentes ou criados e os objetos ou valores são descritos pelos valores literais retirados da base de dados ou por URIs que representam *links* RDF de outros recursos. Foram adicionados também metadados usando termos do vocabulário DC (*Dublin Core*). O *Dublin Core* é um esquema de metadados para descrição de autoria de objetos digitais, tais como vídeos, sons, imagens, textos e *sites* da web.

Em seguida, foram definidos os *links* RDF para relacionar os recursos do projeto com outras fontes de dados. Para isso, foram utilizadas propriedades como *owl:sameAs*, *foaf:homepage*, *foaf:page*, *foaf:primaryTopic*, *rdfs:seeAlso*, *rdf:type* e *skos:subject*. Além disso, algumas informações, como dados geográficos e de ocupação, foram apresentadas como recursos para outras fontes de dados. Dessa forma, *links* RDF foram gerados com os seguintes *data sets*: *DBPedia*, uma versão da *Wikipedia* com dados ligados; *GeoNames*, que possui informações sobre lugares e características geográficas; *Freebase*, uma base de dados com informa-

2 <http://purl.org/vocab/bio/0.1/>

3 <http://models.okkam.org/ENS-core-vocabulary>

4 <http://www.w3.org/2006/vcard/ns>

5 <http://dbpedia.org/property/>

6 <http://www.rdfabout.com/rdf/schema/politico/>

7 <http://purl.org/ontomedia/ext/common/being>

8 <http://motools.sourceforge.net/timeline/timeline.html>

9 <http://purl.org/dc/terms/>

10 <http://www.purl.org/net/rdf-money/>

11 <http://rdf.geospecies.org/ont/geospecies>

12 <http://purl.org/NET/c4dm/event.owl>

13 <http://www.w3.org/2004/02/skos/core>

14 <http://purl.org/ontology/bibo/>

ções sobre diversas entidades do mundo; The World Factbook, publicações da CIA (*Central Intelligence Agency*) sobre os países do mundo; UMBEL (*Upper Mapping and Binding Exchange Layer*), que relaciona o conteúdo e os dados de diferentes entidades; e YAGO, uma ontologia com informações disponibilizadas como dados ligados.

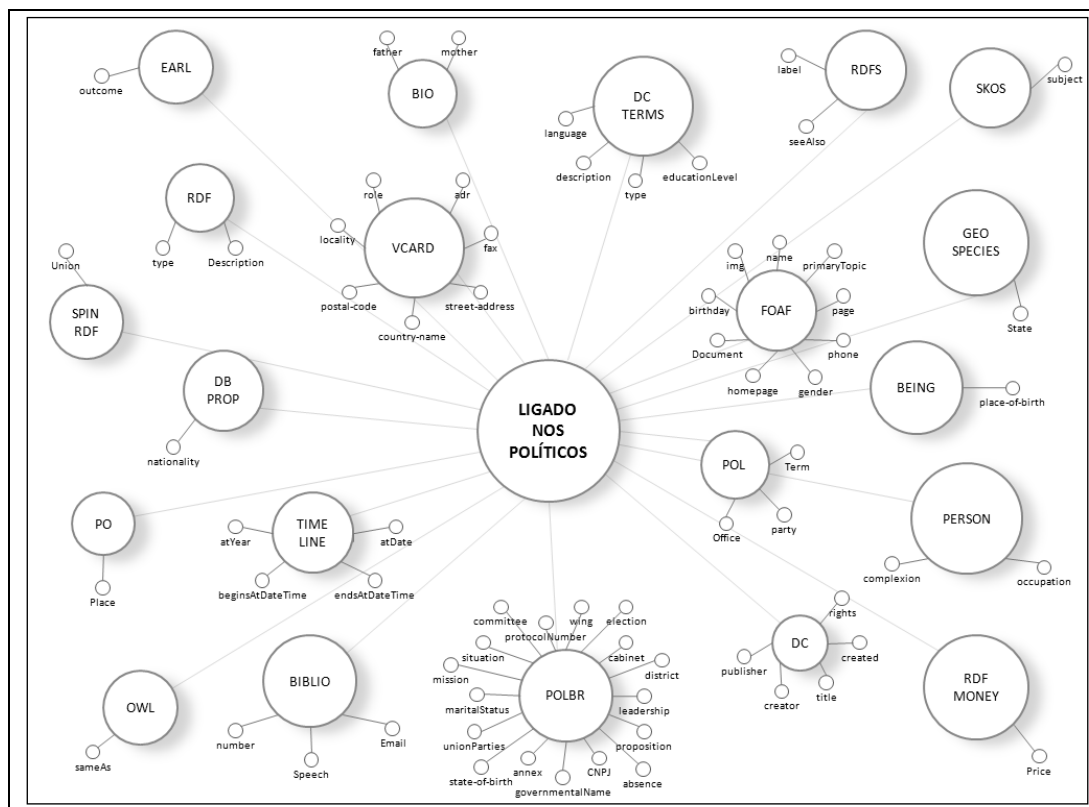


Figura 3. Diagrama de vocabulários e propriedades do projeto

Fonte: dados do trabalho.

A fim de facilitar a descoberta dos dados por máquinas e humanos, o *data set* foi adicionado na *Wiki ESW* e foi realizado um *sitemap* para indicar onde o RDF está localizado. Para testar se os dados podem ser acessados corretamente e se eles podem seguir *links* RDF, foi utilizado o *The Tabulator Extension*, uma extensão para o navegador Firefox que provê uma interface para dados ligados baseada no navegador *The Tabulator*. Também foi utilizado o serviço de validação RDF do W3C para certificar que são fornecidos documentos RDF/XML válidos.

Em seguida, foi criada uma representação HTML para a visualização e consulta dos dados para usuários leigos. A página inicial fornece uma descrição geral do *site*, o número de políticos cadastrados, *links* para outras páginas do *site* e, principalmente, mecanismos de busca para o usuário encontrar os políticos desejados de acordo com diferentes critérios, como nome, situação, cargo, estado, partido e sexo. O usuário pode utilizar mais de um critério para realizar a busca. Diferentes gráficos tam-

bém foram gerados para conferir novos conhecimentos sobre os dados coletados.

Há um *link* para cada recurso não-informacional com o seu respectivo URI. Ao selecionar o político, é exibida a representação adequada de acordo com a requisição do cliente. Os dados do político selecionado são extraídos da base de dados e exibidos dinamicamente nesta página. Para melhorar a navegabilidade do *site*, é possível selecionar certos dados apresentados para buscar políticos que possuem as mesmas características. É apresentado também ao fim da página um *link* para a visualização dos dados no formato RDF/XML.

Com o *data set* de dados de políticos brasileiros *online*, é possível extrair os dados publicados via RDF, realizar consultas em cima desses dados, ligar esses dados com dados de outros *data sets* e desenvolver diferentes aplicações *Web*.

Analisando o projeto de acordo com os aspectos apresentados anteriormente sobre dados governamentais abertos, podemos dizer que os dados publicados atendem de maneira geral os princípios básicos que determinam essa prática: são completos, pois todos os dados coletados estão disponíveis sem limitações de privacidade, segurança ou privilégios de acesso; são primários, pois possuem o maior nível possível de granularidade sem modificações; são atuais, pois são dados recentes de políticos brasileiros, embora esse princípio deva ser avaliado também em longo prazo; são acessíveis e disponibilizados para qualquer usuário da *Web*; são processáveis por máquinas, estruturados usando o modelo RDF; são não discriminatórios, pois não há necessidade de cadastro para acessar os dados; são não proprietários, não havendo uma entidade que detém o controle exclusivo dos dados; e possuem licenças livres.

No que diz respeito aos dados ligados, o projeto também atende aos princípios básicos, usando URIs HTTP para nomear os recursos, permitindo que as pessoas possam acessá-los, fornecendo informações úteis utilizando o padrão RDF e incluindo links para outros URIs para que os usuários possam encontrar mais recursos.

5 CONCLUSÕES

Este trabalho apresentou o desenvolvimento do projeto "Ligado nos Políticos", que concentra informações de políticos brasileiros e disponibiliza as informações em formato de dados ligados, facilitando a reutilização dos dados por outros desenvolvedores. Este é o primeiro *data set* com informações de políticos brasileiros em formato de dados ligados e, até o momento, o único *data set* brasileiro cadastrado na nuvem de dados do projeto *Linking Open Data* da W3C.

Dados governamentais publicados na *Web*, por si só, já possuem um grande valor para a população, pois contribuem para uma maior transparência de informações. Mas a disponibilização dessas informações em

formatos abertos e acessíveis permite que sejam reutilizadas e combinadas com informações de outras fontes para produzir novos significados sobre o desempenho do governo.

Aliar a publicação de dados governamentais abertos às práticas de dados ligados é ainda mais importante, pois proporciona um mecanismo de acesso único e padronizado, permitindo que os dados sejam legíveis por máquinas, facilitando a descoberta e o consumo dos dados, permitindo que eles sejam ligados a outros conjuntos de dados, aumentando o valor e a utilidade dos dados e abrindo possibilidades de aplicações *Web* mais inteligentes.

A publicação de dados governamentais abertos e dados ligados vem crescendo nos últimos anos. Ainda assim, muito ainda deve ser feito para evoluir a *Web* de documentos para uma *Web* de dados e garantir que esses dados sejam abertos e acessíveis para todos, principalmente no que diz respeito a dados governamentais brasileiros.

Atualmente, a publicação de dados governamentais abertos é maior em países como os Estados Unidos e o Reino Unido. Dessa forma, é preciso estender esta prática para os demais países e garantir que mais dados abertos sejam publicados pelos governos. Ao mesmo tempo, mais pessoas e organizações devem publicar dados governamentais por conta própria, contribuindo para o aumento da transparência mesmo nos casos em que os próprios políticos demonstrem relutância.

É importante ressaltar que as publicações devem ir além da esfera de dados de políticos, abrangendo as diferentes áreas da administração pública como saúde, educação, transporte, economia, entre outras. Dessa forma será possível criar mais aplicações, *mashups* e visualizações para oferecer informações úteis aos cidadãos.

O Brasil oferece diversos dados governamentais publicamente, mas é preciso aumentar as iniciativas de dados abertos. Baseando-se nos exemplos bem sucedidos de outros países, devem ser elaborados catálogos ou portais para servir como um ponto único de acesso a dados públicos. São necessárias também mais iniciativas no sentido de extrair os dados já disponíveis e torná-los abertos e reutilizáveis, além da realização de novas aplicações com base nesses dados. Assim, a transparência do governo brasileiro seria aumentada, trazendo impactos sócio-econômicos positivos e melhorando os índices do país em pesquisas que dizem respeito à qualidade do Governo Eletrônico.

No que diz respeito aos dados ligados, existe um limitado número de *data sets* publicados se comparados à quantidade de documentos (X)HTML existentes na *Web* atual. É preciso garantir que mais governos, organizações e pessoas publiquem dados ligados, de forma a aumentar o número de *data sets* interligados. É preciso também melhorar o apoio a infraestrutura técnica para a publicação de dados ligados. Publicações que tratem do tema de forma mais específica devem ser elaboradas de forma a auxiliar desenvolvedores neste processo. Torna-se necessário também

aumentar o número de serviços que auxiliem na geração de *links* RDF, especialmente para o caso onde não existem identificadores comuns entre os *data sets*.

Quanto ao projeto "Ligado nos Políticos", algumas limitações são encontradas, principalmente quanto à abrangência das informações. No *data set* são disponibilizados somente dados de políticos da esfera federal. Além disso, a quantidade de dados relacionados a alguns políticos é muito pequena se comparada a de outros políticos cadastrados. Essas limitações, contudo, são também limitação das fontes de dados utilizadas que não permitem acesso aos dados brutos e disponibilizam somente parte das informações. Como trabalho futuro, pretendemos disponibilizar dados sobre políticos das esferas municipal e estadual. Além disso, utilizar o mesmo processo apresentado nesse artigo para criação de *data sets* de dados ligados com outras informações governamentais, como gastos públicos, editais, projetos de lei etc.

Da mesma forma como o Governo Eletrônico revolucionou a relação entre os cidadãos e o governo, Dados Governamentais Abertos têm o potencial de aprimorar e estreitar ainda mais essa relação. E assim como a *Web* provocou uma revolução na publicação e no consumo de documentos, dados ligados têm o potencial para revolucionar a forma como os dados são acessados e utilizados. Se os desafios ainda existentes forem adequadamente tratados, essas práticas permitirão uma revolução na forma como a *Web* é utilizada atualmente para a comunicação entre governos e cidadãos, ampliando a transparência dos atos públicos a níveis até recentemente inimagináveis.

REFERÊNCIAS

AGUNE, R. M.; GREGORIO FILHO, A. S. ; BOLLIGER, S. P. Governo aberto SP: disponibilização de bases de dados e informações em formato aberto. In: Congresso Consad de Gestão Pública, 3., *Anais...* 2009. Disponível em: http://www.escoladegoverno.pr.gov.br/arquivos/File/Material_%20CONSAD/paineis_III_congresso_consad/painel_13/governo_aberto_sp_disponibilizacao_de_bases_de_dados_e_informacoes_em_formato_aberto.pdf. Acesso em: dezembro de 2010.

BENNETT, D.; HARVEY, A. Publishing open government data. W3C Working Draft. Setembro, 2009. Disponível em: <http://www.w3.org/TR/gov-data/>. Acesso em: dezembro de 2010.

BERNERS-LEE, T. Linked data. Design Issues. Julho, 2006. Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: dezembro de 2010.

BERNERS-LEE, T. Putting government data online. Design Issues. Junho, 2009. Disponível em: <http://www.w3.org/DesignIssues/GovData.html>. Acesso em: dezembro de 2010.

BIZER, C.; CYGANIAK, R.; HEATH, T. How to publish linked data on the web. Julho, 2007. Disponível em: <http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial>. Acesso em: dezembro de 2010.

BIZER, C.; HEATH, T.; IDEHEN, K.; BERNERS-LEE, T. Linked data on the web. In: LDOW 2008 - International World Wide Web Conference. 17., Beijing, *Proceedings...* Abril, 2008. Disponível em: <http://events.linkeddata.org/ldow2008/papers/00-bizer-heath-ldow2008-intro.pdf>. Acesso em: dezembro de 2010.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data - the story so far. Preprint to the special issue on linked data. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 2009. Disponível em: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>. Acesso em: dezembro de 2010.

CYGANIAK, R.; JENTZSCH, A. The linking open data cloud diagram. Setembro, 2010. Disponível em: <http://richard.cyganiak.de/2007/10/lod/>. Acesso em: dezembro de 2010.

DINIZ, V. Como conseguir dados governamentais abertos. In: Congresso Consad de Gestão Pública. 3., *Anais...* 2009. Disponível em: http://www.consad.org.br/sites/1500/1504/000_01870.pdf. Acesso em: dezembro de 2010.

ELMAGARMID, A. K.; IPEIROTIS, P. G.; VERYKIOS, V. S. Duplicate record detection: a survey. Janeiro, 2007. Disponível em: <http://www.cs.purdue.edu/homes/ake/pub/survey2.pdf>. Acesso em: janeiro de 2011.

GI PARA E-GOV - GRUPO DE INTERESSE EM GOVERNO ELETRÔNICO DO W3C. Melhorando o acesso ao governo com o melhor uso da web. Comitê Gestor da Internet no Brasil. 1. ed. São Paulo, 2009. Disponível em: <http://www.w3c.br/divulgacao/pdf/gov-web.pdf>. Acesso em: dezembro de 2010.

GOVERNO ABERTO, Decreto número 55.559 de 12 de março de 2010 para instituição do Portal Governo Aberto de São Paulo, 2010. Disponível em <http://www.governoaberto.sp.gov.br/view/decreto.php>. Acesso em: setembro de 2011.

HEATH, T.; HAUSENBLAS, M.; BIZER, C.; CYGANIAK, R.; HARTIG, O. How to publish linked data on the web: half-day tutorial. In: ISWC2008. Karlsruhe, Alemanha. *Proceedings...* Outubro, 2008. Disponível em: <http://events.linkeddata.org/iswc2008tutorial/how-to-publish-linked-data-iswc2008-slides.pdf>. Acesso em: dezembro de 2010.

HEATH, T., BIZER, C., Linked data: evolving the web into a global data space, synthesis lectures on the semantic web. *Theory and Technology*, v. 1, n. 1, p. 1-136. Morgan & Claypool, 2011.

KLYNE, G.; CARROLL, J. J. *Resource Description Framework (RDF): concepts and abstract syntax*. W3C Recommendation2004. Disponível em

<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>. Acesso em: setembro de 2011.

OPENGOVDATA.ORG. Open government data principles. California. Dezembro, 2007. Disponível em: http://resource.org/8_principles.html. Acesso em: dezembro de 2010.

SHERIDAN, J.; TENNISON, J. Linking UK government data. LDOW. Raleigh, Carolina do Norte. Abril, 2010. Disponível em: http://events.linkeddata.org/ldow2010/papers/ldow2010_paper14.pdf. Acesso em: dezembro de 2010.

THACKER, Comunidade Transparência Hacker, 2011. Disponível em <http://thacker.com.br/node/>. Acesso em: setembro de 2011.

UNITED NATIONS. United Nations e-government survey 2010. Nova Iorque: UN Publishing Section, 2010. Disponível em: <http://unpan1.un.org/intradoc/groups/public/documents/un/unpan038851.pdf>. Acesso em: dezembro de 2010.

W3C. People: Tim-Berners-Lee biography, 2011. Disponível em <http://www.w3.org/People/Berners-Lee/>. Acesso em: setembro de 2011.

W3C ESCRITÓRIO BRASIL. O governo de inovação na Copa 2014: uso de redes sociais e dados abertos. In: Seminário de Inovação em Governo Eletrônico. 2., Porto Alegre, Rio Grande do Sul. *Anais...* Setembro, 2010. Disponível em: http://www.procergs.rs.gov.br/uploads/1285856001W3C_Seminario_Inovacao_eGov_POA_17092010.pdf. Acesso em: dezembro de 2010.

W3C SWEO Community Project. Linking open data on the semantic web – data sets. Dezembro, 2010. Disponível em: <http://esw.w3.org/TaskForces/CommunityProjects/LinkingOpenData/DataSets>. Acesso em: janeiro de 2011.