Publicação de Dados Ligados de Políticos Brasileiros na Web

Lucas de Ramos Araújo

JUIZ DE FORA
DEZEMBRO, 2010

Publicação de Dados Ligados de Políticos Brasileiros na Web

LUCAS DE RAMOS ARAÚJO

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharel em Ciência da Computação

Orientador: Jairo Francisco de Souza

JUIZ DE FORA
DEZEMBRO, 2010

PUBLICAÇÃO DE DADOS LIGADOS DE POLÍTICOS BRASILEIROS NA WEB

Lucas de Ramos Araújo

MONOGRAFIA SUBMETIDADA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA OBTENÇÃO DO GRAU DE BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

_____

Jairo Francisco de Souza, orientador.
MSc em Engenharia de Sistemas e Computação COPPE/UFRJ

_____

Eduardo Barrére
DSc em Engenharia de Sistemas e Computação COPPE/UFRJ

_____

Alessandreia Marta de Oliveira
MSc em Engenharia de Sistemas e Computação COPPE/UFRJ

JUIZ DE FORA, MG – BRASIL
DEZEMBRO, 2010

Agradecimentos

Agradeço à minha família, pelo amor incondicional.

Agradeço ao meu orientador Jairo, pelo exemplo de sabedoria e dedicação.

Agradeço a todos os professores do curso, pelos conhecimentos compartilhados.

Agradeço aos amigos, pelo apoio em todos os momentos.

Agradeço a Fernanda, pela companhia constante.

Resumo

Desde o seu surgimento, a *Web* vem sofrendo constantes evoluções a fim de se aprimorar cada vez mais como meio de colaboração, interação, comunicação global e compartilhamento de informações. A próxima evolução na *Web*, a *Web* Semântica, se propõe a estender os princípios da Web de documentos para uma Web de dados. Dentro da *Web* Semântica, a utilização de Dados Ligados vem crescendo muito nos últimos anos, permitindo o desenvolvimento de aplicações melhores e mais inteligentes. Ao mesmo tempo, Dados Governamentais Abertos estão sendo cada vez mais publicados na *Web*, contribuindo para a transparência e a reutilização dos dados. Neste contexto, este trabalho tem como proposta publicar Dados Governamentais Abertos utilizando as práticas de Dados Ligados, através da criação de um conjunto de dados de políticos brasileiros com informações coletadas de diferentes fontes, contribuindo assim com a nova *Web* de dados.

Palavras-chave: Dados Ligados. Dados Governamentais Abertos. Web Semântica. Governo Eletrônico. Conjuntos de Dados. Transparência.

Abstract

Since its inception, the Web has undergone constant evolution in order to improve itself as a means of collaboration, interaction, global communication and information sharing. Its next evolution, the Semantic Web, aims to extend the principles of the Web documents to a Web of data. Within the Semantic Web, the use of Linked Data has been increasing in recent years, enabling the development of better and smarter applications. At the same time, Open Government Data are increasingly being published on the Web, contributing to the transparency and the reusability of data. 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, thus contributing to the new Web of data.

Keywords: Linked Data. Open Government Data. Semantic Web. Electronic Government. Data Sets. Transparency.

Sumário

Lista de Siglas e Abreviações

API	*Application Programming Interface*
CC	*Creative Commons*
CGI	Comitê Gestor da Internet no Brasil
CONIP	Congresso de Inovação e Informática Na Gestão Pública
CSV	*Comma Separated-Values*
DBPPROP	*DBPedia Property*
DC	*Dublin Core*
DCTERMS	*Dublin Core Terms*
DOAP	*Description of a Project*
e-Gov IG Group	*e-Government Interest Group*
FOAF	*Friend-of-a-Friend*
GATI	Grupo de Apoio Técnico à Inovação
GEO	*GeoNames*
GI para e-Gov	Grupo de Interesse em Governo Eletrônico
GPS	*Global Positioning System*
HTML	*Hypertext Markup Language*
HTTP	*Hypertext Transfer Protocol*
INPE	Instituto Nacional de Pesquisas Espaciais
LOD	*Linking Open Data*
OGDI	*The Open Government Data Initiative*
OKFN	*The Open Knowledge Foundation*
ONG	Organização Não Governamental
OWL	*Web Ontology Language*
PHP	*Hypertext Preprocessor*
POL	*Politico*
PLC	Projeto de Lei da Câmara
RDF	*Resource Description Framework*
RDFa	*RDF – in – attributes*
RDFS	*Resource Description Framework Schema*
REST	*Representational State Transfer*

RSS	*Really Simple Syndication* (RSS 2.0)
SEADE	Fundação Sistema Estadual de Análise de Dados
SQL	*Structured Query Language*
SPARQL	*SPARQL Protocol and RDF Query Language*
SWEO	*Semantic Web Education and Outreach Group*
TIC	Tecnologia da Informação e de Comunicação
TSE	Tribunal Superior Eleitoral
UMBEL	*Upper Mapping and Binding Exchange Layer*
URI	*Uniform Resource Identifier*
URL	*Uniform Resource Locator*
W3C	*World Wide Web Consortium*
XHTML	*Extensible Hypertext Markup Language*
XML	*Extensible Markup Language*
XPath	*XML Path Language*
XSLT	*Extensible Stylesheet Language for Transformation*

Lista de Figuras

Lista de Tabelas

Lista de Listagens

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 eficiente, democrática e transparente.

Dentro deste contexto, o termo Dados Governamentais Abertos surgiu no intuito de ampliar essa relação, promovendo a disponibilização das informações governamentais em formato aberto e acessível de tal modo que possam ser reutilizadas e misturadas com informações de outras fontes, gerando novos significados (W3C Escritório Brasil, 2010).

Ao mesmo tempo, a *Web* vem se aprimorando cada vez mais como meio de colaboração, interação, comunicação global e compartilhamento de informações. Sua próxima evolução, a *Web* Semântica, busca representar os dados com significado para torná-los legíveis por máquinas, abrindo possibilidades de aplicações *Web* melhores e mais inteligentes (MACMANUS, 2010).

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 aumentar o valor e a utilidade desses dados (BIZER *et al.*, 2009).

1.1. Motivação

Dados governamentais publicados na *Web* aumentam a consciência dos cidadãos das funções do governo permitindo uma maior responsabilidade, contribuem com informações valiosas sobre o mundo e permitem que o governo, o país e o mundo funcionem com mais eficiência (BERNERS-LEE, 2008).

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 e sua utilização por máquinas. Para bem aproveitar o potencial representado pelo acervo de informações do governo, eles precisam ser disponibilizados em formato padronizado, aberto e acessível.

Ao mesmo tempo, a utilização de Dados Ligados vem crescendo muito nos últimos anos, sendo fortemente apoiada pelo W3C e por Tim Berners-Lee, considerado o principal inventor da *Web*. Existem diversas maneiras de publicar dados governamentais abertos, mas segundo Berners-Lee (2008) os objetivos esperados ao publicar dados governamentais são melhores alcançadas usando Dados Ligados.

O projeto *Linking Open Data* (LOD) tem como objetivo promover o compartilhamento livre de dados ligados na Web, através da publicação de vários *data sets* (conjuntos de dados) abertos e do estabelecimento de ligações entre eles, criando uma nuvem de dados ligados, denominada LOD *Cloud* (BIZER *et al.*, 2008).

Hoje em dia, há um movimento cada vez maior de governos, organizações e pessoas publicando dados governamentais abertos, inclusive utilizando as práticas de Dados Ligados. Porém, vários desafios devem ser superados para que a *Web* seja utilizada como um grande banco de dados global.

O Brasil, por exemplo, tem uma boa oferta de dados em todas as esferas e poderes, mas existem poucas iniciativas do governo que se propõem a dar acesso à base integral estruturada e com linguagem aberta (AGUNE *et al.*, 2009). Por esse motivo, estão surgindo iniciativas no sentido de extrais os dados, torná-los abertos e conferir novos valores a eles através de diferentes aplicações. Essas iniciativas, porém, ainda não utilizam necessariamente as práticas de Dados Ligados.

É de suma importância repensar o Governo Eletrônico no Brasil. 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 infra-estrutura de telecomunicações. O relatório destaca ainda iniciativas brasileiras de dados abertos que devem ser seguidas.

1.2. Objetivos

Dentro do contexto apresentado, pode ser percebida a importância da publicação de Dados Governamentais Abertos e a relevância das práticas de Dados Ligados na *Web* atual.

Dessa forma, este trabalho tem como objetivo unir ambos os conceitos com a publicação de Dados Ligados de políticos brasileiros na *Web*, através da criação de um *data set* com informações coletadas de diferentes fontes para ser incluído na LOD *Cloud*, 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, de forma a torná-los legíveis tanto por máquinas quanto por humanos.

Busca-se com esse trabalho explorar e discutir os principais conceitos e questões que regem o funcionamento de Dados Ligados e Dados Governamentais Abertos.

Esse trabalho tem como objetivo também difundir as principais práticas que devem ser levadas em consideração na publicação Dados Ligados e de Dados Governamentais Abertos na *Web*, bem como as principais tecnologias e ferramentas que podem ser utilizadas para auxiliar nesse processo.

1.3. Organização do Trabalho

O restante deste trabalho está estruturado da seguinte maneira: O Capítulo 2 aborda o tema Dados Governamentais Abertos, trazendo os seus principais conceitos, benefícios, tecnologias, práticas de publicação e desafios, bem como um panorama geral de sua utilização no Brasil e no mundo. O Capítulo 3 explora o tema Dados Ligados, apresentando os seus principais conceitos, benefícios, tecnologias, desafios e uma visão mais detalhada das práticas de publicação. O Capítulo 4 apresenta as informações referentes ao projeto proposto, além de uma avaliação segundo os conceitos previamente apresentados. Por fim, o Capítulo 5 traz as conclusões obtidas com a realização do trabalho juntamente com propostas de trabalhos futuros.

2. Dados Governamentais Abertos

Neste capítulo são tratadas as principais questões relacionadas a Dados Governamentais Abertos.

A seção 2.1 define o termo em diversos aspectos. A seção 2.2 apresenta os seus princípios fundamentais. A seção 2.3 aborda os benefícios que motivam a sua utilização. A seção 2.4 aponta as principais tecnologias que apóiam a prática. A seção 2.5 apresenta as principais orientações para a publicação de Dados Governamentais Abertos. A seção 2.6 traz uma visão geral de sua utilização no mundo. A seção 2.7 traz uma visão geral de sua utilização no Brasil. A seção 2.8 apresenta os principais desafios na área. Por fim, a seção 2.9 traz as considerações finais.

2.1. Definição

Segundo o W3C Escritório Brasil (2010), Dados Governamentais Abertos podem ser definidos como "a disponibilização de informações governamentais representadas em formato aberto e acessível de tal modo que possam ser reutilizadas, misturadas com informações de outras fontes, gerando novos significados."

O Governo Eletrônico, também denominado *e-gov*, tem como princípio o uso das Tecnologias da Informação e da Comunicação (TICs) para facilitar o acesso aos serviços públicos, permitir ao grande público o acesso à informação, tornar o governo mais transparente para o cidadão e promover maior eficiência e maior efetividade governamental (AIRES, 2006).

Segundo o Grupo de Interesse em Governo Eletrônico – GI para e-Gov (2009) do W3C Escritório Brasil, criar um Governo Eletrônico exige abertura, transparência, colaboração e conhecimento para aproveitar as vantagens da *Web*. 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 o potencial representado pelo acervo de informações das organizações.

Muitos dados governamentais estão disponíveis na *Web*, mas estas informações na maioria das vezes são oferecidas em formatos proprietários, que exigem que o consumidor em potencial tenha o software ou as ferramentas específicas para acessá-las

ou então apenas em formatos humanamente legíveis, sem seguir padrões e limitando seu uso por máquinas. Além disso, o acesso normalmente é feito de forma parcial e fragmentada, e por caminhos pouco transparentes (DINIZ, 2009; AGUNE *et al.*, 2009).

Na maioria das situações, isso significa que o consumidor só tem acesso aos dados do modo como o produtor imagina que eles devem ser acessados, não tendo acesso aos dados brutos e por meio de padrões abertos. Isto impede e dificulta que o interessado possa trabalhar, analisar, cruzar e integrar os dados e informações segundo foco e interesses próprios (DINIZ, 2009; AGUNE *et al.*, 2009).

A disponibilização de Dados Governamentais Abertos permite que os usuários possam facilmente encontrar, acessar, entender e utilizar os dados públicos. A representação dos dados de uma maneira que as pessoas possam reutilizá-los é um dos passos mais relevantes para a caracterização dos dados como Dados Governamentais Abertos (DINIZ, 2009).

2.2. Princípios

O *Open Government Working Group* (OPENGOVDATA.ORG, 2007), elaborou os 8 Princípios dos Dados Governamentais Abertos. Eles devem ser:

- **Completos:** Todos os dados estão disponíveis e não limitados. Um dado público é o dado que não está sujeito a limitações válidas de privacidade, segurança ou privilégios de acesso;
- **Primários:** Os dados são coletados na fonte, com o maior nível possível de granularidade, sem agregação ou modificação;
- **Atuais:** Os dados são publicados tão rapidamente quanto necessário para preservar o seu valor;
- **Acessíveis:** Os dados são disponibilizados para o maior número possível de usuários e para o maior número possível de finalidades;
- **Processáveis por máquinas:** Os dados são razoavelmente estruturados para permitir processamento automatizado;
- **Não-discriminatórios:** Os dados são disponíveis para todos, sem necessidade de cadastro;
- **Não-proprietários:** Os dados são disponibilizados em um formato sobre o qual nenhuma entidade tem controle exclusivo;

- **Licenças livres:** Os dados não estão sujeitos a nenhuma regulação de direitos autorais, patentes, propriedade intelectual ou segredo industrial. Restrições sensatas relacionadas à privacidade, segurança e privilégios de acesso podem ser permitidas.

Com o mesmo propósito, Eaves (2009) elaborou as três leis dos dados governamentais abertos: (1) se ele não pode ser encontrado na *Web* e indexado, ele não existe; (2) se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser utilizado; (3) se qualquer dispositivo legal não permitir que ele seja reutilizado, ele não é útil.

2.3. Benefícios

De uma maneira geral, a publicação de dados governamentais aumenta a consciência dos cidadãos das funções do governo permitindo uma maior responsabilidade; contribuem com informações valiosas sobre o mundo e permitem que o governo, o país e o mundo funcionem com mais eficiência (BERNERS-LEE, 2009).

De acordo com Bennet e Harvey (2009), Sheridan e Tennison (2010) e o GI para e-Gov (2009), dados governamentais publicados em formatos padronizados, abertos e acessíveis na *Web* podem trazer diversos benefícios, apresentados de forma resumida a seguir:

- **Reutilização:** permitem a mistura de dados de vários aplicativos ou fontes de dados (*mashup*) de maneiras novas, imprevistas e imaginativas, aumentando muito o valor dos dados por sua combinação e produzindo novos conhecimentos e serviços;
- **Inclusão:** permitem que qualquer pessoa use numerosas ferramentas de software para adaptá-los às suas necessidades. O conjunto de dados pode ser transformado e colocado à disposição de vários equipamentos, inclusive para tecnologias usadas por pessoas com deficiências;
- **Transparência:** as partes interessadas podem usar as informações da maneira mais adequada ao seu propósito, fornecendo aos cidadãos uma idéia melhor do trabalho do governo e adaptando-o às suas necessidades específicas;

- **Responsabilidade:** podem oferecer vários pontos de vista sobre o desempenho do governo ao tentar atingir suas metas em políticas públicas, permitindo o controle social e oficial sobre a atuação adequada da administração;
- **Melhoria nas buscas:** melhoram as pesquisas na *Web*, ajudando ainda mais os consumidores a encontrarem as informações que precisam;
- **Integração:** permitem que outros órgãos e entidades misturem, melhorem e partilhem essas informações, o que produz uma grande melhoria na integração de dados entre sistemas díspares, e o florescimento de novos serviços;
- **Participação:** possibilitam aos cidadãos a crescente capacidade para participar e influenciar nas decisões político-administrativas que lhe digam respeito;
- **Colaboração:** garantem a constante colaboração entre os diversos instrumentos e entidades da Web;
- **Crescimento Econômico:** apoiam o crescimento econômico, estimulando novos produtos e serviços baseados em dados;
- **Inovação:** incentivam à criação de uma cultura voltada para a importância da inovação e da geração e compartilhamento de conhecimento e informação na gestão pública, estimulando e promovendo inovações no governo, principalmente no seu relacionamento com a sociedade;
- **Eficiência:** aumentam a velocidade e qualidade dos serviços públicos e do compartilhamento de informações e conhecimentos para os cidadãos.

2.4. Tecnologias

A maneira mais rápida e fácil de tornar os dados disponíveis na *Web* é publicá-los em sua forma bruta. No entanto, os dados devem ser bem estruturados. A estrutura permite que outros façam uso automatizado dos dados com sucesso. Formatos que só permitem a visualização dos dados não são úteis e devem ser evitados (BENNETT e HARVEY, 2009).

O W3C tem produzido uma série de tecnologias que combinadas com outras existentes viabilizam a disponibilização de dados abertos e oferecem ótima manipulação e conjuntos de ferramentas padronizadas (DINIZ, 2009).

As principais tecnologias utilizadas para a publicação de dados governamentais abertos são apresentadas nas seções seguintes.

2.4.1. Arquivos CSV

O CSV (*Comma Separated-Values*) é um formato de arquivo que armazena dados tabulares. Ele separa campos (colunas) com vírgulas e registros (linhas) com a quebra de linha. Tem a vantagem de ser comum a todas as plataformas de computador.

É a forma mais simples de publicar dados governamentais abertos. Programadores podem usar os dados do arquivo e importá-los em uma base de dados e não-programadores podem abrir o arquivo em planilhas eletrônicas e gerar gráficos (BENNETT e HARVEY, 2009).

2.4.2. Informações RSS/Atom

RSS (*Really Simple Syndication*) é um formato para agregar conteúdo na *Web*, baseado em XML (*Extensible Markup Language*) (RSS 2.0 SPECIFICATION, 2003). Atom é um formato de documento baseado em XML que descreve listas de informações relacionadas (NOTTINGHAM e SAYRE, 2005). Esses formatos são usados geralmente para compartilhar novidades ou textos completos através dos denominados *feeds*, compostos por um número de itens, conhecidas como entradas, cada uma com um conjunto extensível de metadados.

Muitos elementos de informações fornecidas por governos são adequados para serem distribuídos como *feeds* de notícias, usando RSS ou Atom, e podem ser usados com um grande número de ferramentas e navegadores. As pessoas podem assinar um conjunto de canais e receber informações atualizadas sobre, por exemplo, notícias do governo, vagas de emprego, concessões ou aquisições. Os usuários precisam apenas de um leitor de *feed* para assinar e ler as informações. O número de *feeds* oferecidos por governos cresce constantemente, e já há milhares deles disponíveis (GI PARA E-GOV, 2009).

2.4.3. Interfaces REST

O estilo arquitetural REST (*Representational State Transfer*) oferece uma arquitetura para criar aplicativos na Web usando HTTP (*Hypertext Transfer Protocol*) e URIs (*Uniform Resource Identifiers*), as duas especificações que definem a interface genérica usada por todas as iterações de componentes na Web (FIELDING, 2000; GI PARA E-GOV, 2009).

Basicamente, ele associa um recurso a um URI que pode ser usado para acessar ou modificar suas informações de acordo com alguns princípios de criação, permitindo que um site possa ser enriquecido com aplicativos que expanda o valor de um recurso disponível (DINIZ, 2009; GI PARA E-GOV, 2009).

Esse modelo é altamente adequado ao desenvolvimento de *mashups* e também pode fornecer dados em formatos brutos abertos, como no site *Seniors Canada Online*[1], que oferece essas interfaces para realizar buscas em suas bases de dados, podendo ser usadas por outros órgãos (GI PARA E-GOV, 2009).

2.4.4. Tecnologias da Web Semântica

A *Web* Semântica oferece um arcabouço comum, que permite que dados sejam compartilhados e reutilizados além dos limites de aplicativos, empreendimentos e comunidades. Existem várias tecnologias que permitem descrever, modelar e pesquisar esses dados (GI PARA E-GOV, 2009).

Segundo Berners-Lee (2008) os objetivos esperados ao publicar dados governamentais são melhores alcançadas usando Dados Ligados. A Web Semântica oferece um alto grau de automação. Ainda que outras tecnologias atuais (serviços na *Web*, REST, etc.) ofereçam esse tipo de automação, é impossível prever todos os cenários de uso dos dados, tornando o seu uso limitado (GI PARA E-GOV, 2009).

Os padrões, tecnologias e recursos da *Web* Semântica serão discutidos com mais detalhes nos capítulos subsequentes, bem como os benefícios trazidos ao utilizá-los na publicação e utilização de Dados Governamentais Abertos.

[1] http://www.seniors.gc.ca/

2.5. Publicação

Segundo Bennett e Harvey (2009), para publicar Dados Governamentais Abertos, é preciso tomar três medidas fundamentais: (1) selecionar que dados serão disponibilizados e identificar quem os controla; (2) representar esses dados de uma maneira que as pessoas possam reutilizá-los; e (3) publicar os dados e divulgar.

Diniz (2009) descreve também uma série orientações para disponibilizar dados de modo aderente às características que definem os dados governamentais abertos, apresentadas de maneira resumida a seguir:

- Publicar inicialmente os dados em sua forma bruta de forma estruturada;
- Criar um catálogo online dos dados brutos com documentação para que as pessoas possam descobrir o que foi postado;
- Usar normas estabelecidas abertas e ferramentas que permitam uma produção e publicação fácil e eficiente;
- Tornar os dados legíveis para pessoas convertendo-os para (X)HTML;
- Deixar as páginas legíveis por máquinas incorporando informações semânticas, metadados e identificadores;
- Usar URIs permanentemente padronizados e/ou de fácil localização;
- Permitir citações eletrônicas sob a forma de *hiperlinks* padronizados;
- Organizar os dados do catálogo usando formatos como o RSS para facilitar e agilizar a divulgação dos conjuntos de dados após sua publicação;
- Utilizar IDs internos para identificar os dados específicos para reutilização por máquinas;
- Criar uma página web com uma descrição clara do conjunto de dados;
- Quando possível, documentar mecanismos de busca e métodos *RESTful* de obtenção dos dados;
- Gerar *links* para outros URIs em seus dados para ajudar na descoberta de recursos relacionados;
- Garantir que os dados são de fácil recuperação e podem ser referenciados pelo tempo que for necessário;
- Integrar novos URIs em conjuntos de dados novos e atualizados, e estruturá-los em conformidade;

- Documentar cuidadosamente as mudanças entre as versões e inserir o número da versão/indicador nos dados;
- Não comprometer a integridade dos dados apenas para criar interfaces chamativas;
- Publicar também os dados brutos separadamente mesmo com a criação de uma interface;
- Publicar todos os dados que já estão disponíveis com o público de outras maneiras;
- Fornecer a documentação completa para permitir a geração automática de dicionários de dados;
- Fornecer serviços de busca, para facilitar a recuperação de documentos e bases de dados;
- Documentar claramente qualquer restrição legal ou regulatória à utilização dos dados.

2.6. Dados Governamentais Abertos no Mundo

Há um movimento global de governos e autoridades locais começando a colocar seus dados na web. Projetos de dados governamentais abertos surgiram em vários países do mundo, como Estados Unidos[2], Reino Unido[3], Austrália[4], Nova Zelândia[5], Noruega[6], Holanda, Suécia, Espanha, Estônia, Áustria, Grécia[7], Canadá e Dinamarca, existindo também um número crescente de iniciativas locais de estados e cidades como Vancouver, Londres e Nova York (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). Além disso, pessoas e organizações vêm publicando dados governamentais por conta própria em vários formatos (BERNERS-LEE, 2009).

Com tantos dados governamentais para trabalhar, desenvolvedores estão criando uma ampla variedade de aplicações, *mashups* e visualizações, especialmente nos

[2] http://data.gov/
[3] http://data.gov.uk/
[4] http://data.australia.gov.au/
[5] http://data.govt.nz/
[6] http://data.norge.no/
[7] http://geodata.gov.gr/

Estados Unidos e no Reino Unido. Estas aplicações oferecem informações muito úteis aos cidadãos, mostrando o potencial da reutilização dos dados governamentais abertos.

O projeto *Where Does My Money Go?*[8] permite aos usuários explorarem dados de gastos públicos do governo do Reino Unido usando diversos mapas, timelines e gráficos. O site *ITO World*[9] trabalha com estatísticas do departamento de transporte do Reino Unido publicados no portal *data.gov.uk* para mostrar informações úteis de transporte. O *Visualizing Community Health Data*[10] traz recursos que permitem o acesso aos dados do departamento americano de Saúde e Serviços Humanos através de tabelas classificáveis, mapas e gráficos de dispersão. O *FixMyStreet*[11] é um *site* no qual moradores do Reino Unido podem relatar problemas na sua vizinhança, como por exemplo pichações ou buracos na rua.

Alguns governos organizam também concursos para descobrir quais são os aplicativos mais procurados, como *Show Us a Better Way*[12] no Reino Unido e o *Apps for Democracy*[13], do Distrito de Columbia nos Estados Unidos (GI para e-Gov, 2009).

No que diz respeito a dados de políticos, existem iniciativas mundiais que trazem dados abertos ou ainda aplicações que através de reutilização de dados fornecem informações muito úteis aos cidadãos.

O *TheyWorkForYou UK*[14] é um *mashup* que traz opiniões e muitos tipos de dados sobre o trabalho dos representantes eleitos no Reino Unido. O *OpenCongress*[15] traz dados governamentais e novidades sobre contas, senadores e representantes do governo dos Estados Unidos. O *GovTrack*[16] permite acompanhar as atividades legislativas, votos e estatísticas de membros do congresso dos Estados Unidos. O *OpenAustralia*[17] apresenta informações e notícias sobre os representantes do Parlamento do governo da Austrália. O site canadense *OpenParlamientCA*[18] publica dados de políticos parlamentares do Canadá e as leis que eles estão propondo.

Existem mundialmente diversas entidades que incentivam a publicação e a utilização de dados governamentais abertos, como o consórcio W3C, principalmente

8 http://www.wheredoesmymoneygo.org/
9 http://itoworld.blogspot.com/
10 http://health.jameyer.com/
11 http://www.fixmystreet.com/
12 http://showusabetterway.com/
13 http://www.appsfordemocracy.org/
14 http://www.theyworkforyou.com
15 http://www.opencongress.org/
16 http://www.govtrack.us/
17 http://www.openaustralia.org/
18 http://openparliament.ca/

através do grupo *e-Government Interest Group* (e-Gov IG Group)[19], a fundação *The Open Knowledge Foundation* (OKFN)[20] e a iniciativa *The Open Government Data Initiative* (OGDI)[21].

Além disso, diversos eventos estão sendo realizados sobre o assunto, como a Conferência para Parlamentares: Transparência na Era Digital (*Conference for Parliamentarians: Transparency in the Digital Era*) e o *Open Government Data* Camp.

2.7. Dados Governamentais Abertos no Brasil

Segundo Agune (2009), o Brasil tem uma boa oferta e dados em todas as esferas e poderes oferecidos publica e gratuitamente, mas na maioria das vezes estes dados estão em formatos fechados ou apenas de exibição. Existem poucas iniciativas do governo que se propõem a dar acesso à base integral estruturada e em linguagem aberta.

Uma dessas iniciativas é o projeto Governo Aberto SP[22], que disponibiliza para a sociedade bases de dados e informações atualizadas do Governo do Estado de São Paulo em caráter aberto e gratuito, estruturado pela Secretaria de Gestão Pública por meio do Grupo de Apoio Técnico à Inovação (GATI), em parceria com a Fundação Sistema Estadual de Análise de Dados (SEADE) e com o apoio institucional e técnico do W3C.

O relatório *United Nations E-Government Survey* 2010 (UNITED NATIONS, 2010) indica a iniciativa do governo paulista como boa prática e caminho a ser seguido, citando que o Estado de São Paulo, no Brasil, está seguindo um caminho semelhante ao governo dos Estados Unidos para a criação de um *site* que irá servir como um ponto único de acesso a dados públicos.

Outra iniciativa do governo é o projeto federal LeXML[23], portal especializado em informação jurídica e legislativa que reúne leis, decretos, acórdãos, súmulas, projetos de leis entre outros documentos das esferas federal, estadual e municipal dos Poderes Executivo, Legislativo e Judiciário de todo o Brasil, no intuito de organizar, integrar e dar acesso às informações disponibilizadas nos diversos portais de órgãos do governo.

Dada a escassez de projetos de Dados Governamentais Abertos, são poucos ainda os projetos que misturam diferentes fontes de informações publicadas para criar

19 http://www.w3.org/2007/eGov/IG/wiki/
20 http://okfn.org/
21 http://ogdisdk.cloudapp.net/
22 http://www.governoaberto.sp.gov.br/
23 http://projeto.lexml.gov.br/

serviços que ofereçam uma nova perspectiva sobre diferentes esferas da administração pública.

Por esse motivo, enquanto o governo não libera os 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 conferir novo valor a eles através de diferentes aplicações.

O Congresso Aberto[24], em fase de teste, visa aumentar a transparência e contribuir para debates acerca do legislativo brasileiro, facilitando o acesso à informação e análises sobre o tema. Utiliza dados oficiais providos por diversos órgãos do governo para gerar um panorama completo da atuação de parlamentares e partidos.

O Parlamento Aberto[25], em fase de estruturação, é um projeto de democracia eletrônica que tem como objetivo tornar mais transparente a atuação do Legislativo, permitindo acompanhar as votações, além de dar ao cidadão comum as ferramentas necessárias para exercer a atividade legislativa, votando nas mesmas leis que os legisladores e propondo suas próprias.

O Legisdados[26] é um projeto que tem como objetivo espelhar os dados de tramitação parlamentar no Brasil, inicialmente da Câmara e Senado, extensível a outras casas legislativas estaduais e municipais.

O Tr3s[27] é um *mashup* com o *Google Maps*[28] para tentar mostrar de forma amigável os dados disponibilizados no site do INPE (Instituto Nacional de Pesquisas Espaciais) sobre o desmatamento.

O SACSP[29] é um sistema interativo de estatísticas e acompanhamento das reclamações de municípios na cidade de São Paulo. A missão do site é ajudar os cidadãos a fiscalizarem o trabalho público em seus bairros usando a plataforma *Web*. Todos os dados disponibilizados vêm do site da Prefeitura da Cidade de São Paulo de forma automatizada.

O site Xerifes do DF[30] mostra em mapa as zonas eleitorais do Distrito Federal, com informações sobre os "xerifes" de cada área, incluindo o número de votos recebidos e o partido político.

---

24 http://www.congressoaberto.com.br/
25 http://trac.meuparlamento.org/
26 http://www.legisdados.org/
27 http://tree.veredas.net/
28 http://maps.google.com/
29 http://sacsp.mamulti.com/
30 http://eleicoes.mamulti.com/

Existem também no Brasil diversos grupos e entidades que promovem os dados governamentais abertos, como o consórcio W3C Escritório Brasil, através do Grupo de Interesse para e-Governo. A comunidade Transparência *Hacker*[31] é um espaço para que profissionais e cidadãos proponham e articulem idéias e projetos que utilizem a tecnologia para fins de interesse da sociedade, trabalhando com Dados Governamentais Abertos e promovendo ações que evidenciam a importância desses dados.

Eventos sobre o Governo Eletrônico, como por exemplo, o CONIP (Congresso de Inovação e Informática Na Gestão Pública), estão colocando o assunto em pauta. Outros eventos estão sendo criados para debater especificamente o tema, como por exemplo, o *Transparency HackDay*.

Dado o crescente interesse civil após exemplos bem sucedidos em outros países, mais iniciativas de dados governamentais abertos deverão ser elaboradas em esferas políticas brasileiras.

No que diz respeito à demanda dos usuários por serviços de Governo Eletrônico, a Pesquisa sobre o uso das Tecnologias da Informação e da Comunicação no Brasil 2009, realizada pelo Comitê Gestor da Internet no Brasil (CGI), indica que 27% dos entrevistados utilizaram algum serviço do Governo Eletrônico (CGI, 2010).

Está em tramitação, com data de votação no congresso indefinida, a chamada Lei de Acesso à Informação Pública (PLC - PROJETO DE LEI DA CÂMARA, Nº 41 de 2010)[32], que regulamentará o direito de acesso de qualquer pessoa a informações de órgãos públicos de interesses particulares ou coletivos, como previsto na Constituição da República no inciso XXXIII do artigo 5º.

2.8. Desafios

O Grupo de Interesse em Governo Eletrônico (2009) evidencia uma série de problemas e desafios enfrentados atualmente na utilização de Dados Governamentais Abertos.

De maneira geral, o fornecimento de Dados Governamentais Abertos não tem recebido recursos humanos ou financeiros, e os órgãos governamentais ainda não consideraram seriamente o uso coordenado de *mashups*.

Os órgãos governamentais têm o desafio de encontrar outros órgãos ou organizações cujos regulamentos ou políticas permitem a troca de informações. Muitas

[31] http://thacker.com.br/

[32] http://www.senado.gov.br/atividade/materia/detalhes.asp?p_cod_mate=96674

vezes os departamentos não consideram que uma de suas missões é oferecer conjuntos de informações de outros órgãos ou fontes diferentes.

Oferecer acesso a dados em formatos abertos transfere o controle e o gerenciamento dos dados para fora do órgão responsável, portanto o órgão não pode mais ter certeza de que os dados mantiveram seu caráter original, e o consumidor final não pode ter certeza se os dados são confiáveis ou não.

Embora algumas das tecnologias e padrões já estejam em uso há muitos anos, pode haver casos nos quais o seu uso causará alguns problemas, ou em que não será possível aplicar a tecnologia da maneira como se pretendia – isto é, existem algumas falhas nos padrões, ou necessidades de novos recursos.

2.9. Considerações finais

No presente capítulo foram apresentados os principais conceitos relacionados a Dados Governamentais Abertos e a sua importância nos dias de hoje. Foram abordadas as principais tecnologias que suportam esta prática, e foram apontados os benefícios da utilização das tecnologias da *Web* Semântica neste contexto.

O próximo capítulo apresenta o conceito de Dados Ligados, uma forma de aplicar as tecnologias da *Web* Semântica para ligar dados de diferentes fontes de uma forma padronizada.

3. Dados Ligados

Neste capítulo são abordados os principais conceitos e princípios técnicos relacionados aos Dados Ligados.

A seção 3.1 define o termo em diferentes aspectos. A seção 3.2 apresenta os seus princípios básicos. A seção 3.3 trata dos benefícios que motivam a sua utilização. A seção 3.4 apresenta as principais tecnologias que apóiam a prática. A seção 3.5 apresenta o projeto *Linking Open Data*. A seção 3.6 descreve algumas aplicações desenvolvidas na área. A seção 3.7 traz as principais práticas necessárias para publicar Dados Ligados na *Web*. A seção 3.8 apresenta os desafios e pesquisas atuais na área. Por fim, a seção 3.9 traz as considerações finais.

3.1. Definição

O termo Dados Ligados se refere a um conjunto de práticas para a publicação e conexão de dados estruturados na *Web* (BIZER *et al.*, 2009). O pressuposto básico por trás de Dados Ligados é que o valor e a utilidade dos dados aumentam se eles estão interligados com outros dados (BIZER *et al.*, 2007).

Dados Ligados é uma parte da *Web* Semântica, em que os dados são representados com significado na *Web*, abrindo muitas possibilidades de aplicações web mais inteligentes (MACMANUS, 2010).

Segundo Berners-Lee (2006), a *Web* Semântica não se refere apenas a colocar dados na *Web*, mas sim fazer ligações de modo que pessoas ou máquinas possam explorar a *Web* de Dados. Com Dados Ligados, a partir de algum dado você pode encontrar outros dados relacionados a ele.

Tecnicamente, Dados Ligados refere-se a dados publicados na *Web* de modo que eles sejam legíveis por máquina, os seus significados sejam explicitamente definidos, eles 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 utilização de Dados Ligados vem crescendo muito nos últimos anos, permitindo uma nova classe de aplicações e a criação de um espaço global de dados (BIZER *et al.*, 2009).

3.2. Princípios

A idéia básica de Dados Ligados foi elaborada por Berners-Lee (2006). Ele definiu os quatro princípios que caracterizam os Dados Ligados e que devem ser aplicados para fazer a *Web* crescer:

1. Usar URIs para nomes de "coisas";
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".

As tecnologias que apóiam a utilização de Dados Ligados serão descritas com mais detalhe na seção 3.4.

3.3. Benefícios

Dados Ligados é uma forma de publicar dados na *Web* que facilita a descoberta e o consumo desses dados, maximiza as suas inter-relações, reduz a redundância, promove a reutilização e permite adicionar valor a esses dados (HEATH, 2009).

Segundo Berners-Lee (2009), os benefícios gerais de Dados Ligados são: eles são acessíveis através de uma variedade ilimitada de aplicações e aplicativos porque são expressos em formatos abertos e não-proprietários; podem ser combinados através de *mashups* com qualquer outro conjunto de Dados Ligados, sendo que nenhum planejamento antecipado é necessário para integrar essas fontes de dados desde que ambos utilizem os padrões de Dados Ligados; é fácil acrescentar mais Dados Ligados aos que já existem, mesmo quando os termos e definições usadas mudam ao longo do tempo.

Com conjuntos de dados conectados na *Web* é naturalmente desejável realizar consultas sobre vários conjuntos de dados de uma vez utilizando SPARQL, permitindo assim uma nova classe de aplicações (ALEXANDER *et al.*, 2009).

Comparados com os dados estruturados acessíveis na *Web* através de APIs (*Application Programming Interfaces*) da *Web* 2.0, Dados Ligados têm a vantagem de proporcionar um mecanismo de acesso único e padronizado, em vez de utilizar diversas

interfaces e formatos. Isso permite que as fontes de dados sejam mais facilmente indexadas pelos mecanismos de busca e permite ligações entre dados de diferentes fontes (BIZER *et al.*, 2007).

3.4. Tecnologias

As principais tecnologias que apóiam os Dados Ligados são: URIs, HTTP e RDF (BIZER *et al.*, 2009).

Além dessas, outras tecnologias da *Web* Semântica são utilizadas para fornecer suporte de diversas formas, por exemplo, na consulta de dados, na definição de vocabulários e na publicação de dados com significado.

As seções seguintes descrevem conceitos importantes sobre as principais tecnologias utilizadas em Dados Ligados.

3.4.1. URIs

Um URI é uma cadeia de caracteres compacta para identificar um recurso físico ou abstrato (BERNERS-LEE, 1998).

Todos os itens de interesse na *Web* são chamados de recursos. Eles são as entidades cujas propriedades e relações queremos descrever com os dados. (BIEZER *et al.*, 2007).

Existem dois tipos de recursos: recursos informacionais e recursos 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. Um único recurso informacional pode ter várias representações, por exemplo, em diferentes formatos, qualidades de resolução ou linguagens (BIEZER *et al.*, 2007).

3.4.2. RDF

O RDF é um modelo padrão para representar dados na Web. Ele amplia a estrutura de *links* da Web usando URIs para indicar a relação entre as entidades. Este modelo

simples permite que dados estruturados e semi-estruturados sejam misturados, expostos e compartilhados entre diferentes aplicações (RDF WORKING GROUP, 2006).

Segundo Breitman (2005), RDF é uma linguagem declarativa que fornece uma maneira padronizada de utilizar XML para representar metadados no formato de sentenças sobre propriedades e relacionamentos entre recursos na *Web*.

Em RDF, a descrição de um recurso é representada como uma série de triplas. As três partes de cada tripla são chamados 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 relacionada ao sujeito. O predicado é um URI de algum vocabulário que indica o tipo de relação que existe entre o sujeito e o objeto (BIZER *et al.*, 2007).

A listagem 3.1 exemplifica a utilização do modelo RDF, descrevendo que o recurso *"http://www.politicos.com/jose"* (sujeito) possui a propriedade *"http://xmnls.com/foaf/0.1/name"* (predicado) descrita pelo valor *"José dos Santos"* (objeto).

```
<?xml version-"1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:foaf="http://xmnls.com/foaf/0.1/">
    <rdf:Description rdf:about="http://www.politicos.com/jose">
        <foaf:name>José dos Santos</foaf:name>
    </rdf:Description>
</rdf:RDF>
```

Listagem 3.1: Exemplo da utilização do modelo RDF

Triplas literais têm como o objeto um valor literal, usado para descrever as propriedades dos recursos, enquanto *links* RDF representam ligações entre dois recursos onde os URIs do sujeito e do objeto identificam os recursos interligados e o URI do predicado define o tipo da ligação (BIZER *et al.*, 2007).

Enquanto as unidades primárias da *Web* de Hipertexto são documentos HTML conectados por *hiperlinks*, Dados Ligados utilizam o modelo RDF para publicar dados na Web e *links* RDF para interligar dados de diferentes fontes (BIZER *et al.*, 2009).

3.4.3. RDFS

A linguagem RDF não fornece mecanismos para descrever as propriedades, nem prevê mecanismos para descrever as relações entre essas propriedades e outros recursos. O papel do RDFS ou RDF-*Schema*, uma linguagem de descrição de vocabulários, é definir classes e propriedades que podem ser utilizadas para essa finalidade (BRICKLEY e GUHA, 2004).

De maneira geral, o RDFS permite que os recursos descritos no documento RDF sejam definidos como instâncias ou subclasses das classes presentes no RDFS; permite definir relacionamentos entre as classes; torna a informação mais fácil de ser entendida por leitores humanos através de comentários e etiquetas; e permite definir termos de restrição como domínio e alcance da propriedade (BREITMAN, 2005).

A listagem 3.2 exemplifica a utilização do RDFS para definir relacionamentos entre diferentes classes.

```
<?xml version-"1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:rdfs="http://www.w3.org/2000/01-rdf-schema#"
         xml:base="http://www.politicos.com/politicos/">
    <rdfs:Class rdf:ID="politico" />
    <rdfs:Class rdf:ID="deputado" />
        <rdfs:subClassOf rdf:resource="#politico" />
    </rdfs:Class>
</rdf:RDF>
```

Listagem 3.2: Exemplo da utilização do modelo RDFS

3.4.4. OWL

A OWL (*Web Ontology Language*) é uma linguagem usada para explicitar os significados dos termos em vocabulários e as relações entre esses termos. Esta representação dos termos e de seus inter-relacionamentos é chamada de ontologia (MCGUINNESS e HARMELEN, 2004).

A OWL fornece vocabulários adicionais para descrever propriedades e classes como conectivos lógicos, negação, disjunção, conjunção, cardinalidade, igualdade, tipos de propriedades mais ricos, características de propriedades e classes enumeradas (BREITMAN, 2005).

A listagem 3.3 mostra um exemplo da utilização da OWL, onde a propriedade *owl:disjointWith* explicita que uma classe não pode compartilhar instâncias com classes que mantêm esse tipo de relacionamento.

```
<owl:Class rdf:ID="Masculino">
    <rdfs:subClassOf rdf:resource="#Sexo" />
    <owl:disjointWith rdf:resource="Feminino" />
</owl:Class>
```

Listagem 3.3: Exemplo da utilização da linguagem OWL (BREITMAN, 2005)

3.4.5. RDFa

RDFa (RDF – *in* – *attributes*) permite usar atributos XHTML para marcar dados legíveis por humanos com indicadores legíveis por máquina para serem interpretados por navegadores e outros programas (ADIDA *et al.*, 2008).

Com a utilização do RDFa é possível incorporar triplas RDF em documentos XHTML, como mostra a listagem 3.4. É possível também extrair essas triplas através de ferramentas específicas, como por exemplo, o *RDFa Distiller and Parser*[33].

```
<html xmlns="http://www.w3.org/1999/xhtml"
      xmlns:biblio="http://example.org/"
      xmlns:dc="http://purl.org/dc/elements/1.1/">
  <head>
    <title>Livros de Marco Pierre White</title>
  </head>
  <body>
    O livro
    <span about="urn:ISBN:0091808189" typeof="biblio:book"
property="dc:title">
      Canteen Cuisine
    </span>
    é uma boa dica de leitura.
  </body>
</html>
```

Listagem 3.4: Exemplo da utilização da marcação RDFa (ADIDA *et al.*, 2008)

[33] http://www.w3.org/2007/08/pyRdfa/

3.4.6. SPARQL

A SPARQL é a linguagem de consulta para RDF, podendo ser usada para expressar consultas através de diferentes fontes de dados (PRUD'HOMMEAUX e SEABORNE, 2008).

Essa linguagem permite que arquivos RDF sejam consultados através de uma linguagem parecida com SQL (*Structured Query Language*). O exemplo da listagem 3.5 mostra uma simples consulta SPARQL para encontrar o título de um livro de um dado grafo de dados.

```
Dados:
Sujeito   -> <http://example.org/book/book1>
Predicado -> <http://purl.org/dc/elements/1.1/title>
Objeto    -> "Tutorial SPARQL "

Consulta:
SELECT ?title
WHERE
{
  <http://example.org/book/book1> <http://purl.org/dc/elements/1.1/title>
  ?title .
}
Resultado da Consulta:
title -> "Tutorial SPARQL "
```

Listagem 3.5: Consulta SPARQL (PRUD'HOMMEAUX e SEABORNE, 2008)

3.4.7. URIs HTTP Desreferenciáveis

Desreferenciar um URI é o processo de acessar um URI na Web a fim de obter as informações sobre o recurso referenciado (BIEZER *et al.*, 2007).

Quando um URI que identifica um recurso informacional é desreferenciado, o servidor do URI gera uma nova representação do estado atual do recurso informacional e a envia para o cliente utilizando o código de resposta HTTP 200 OK (BIEZER *et al.*, 2007).

Recursos não-informacionais não podem ser desreferenciados diretamente. Existem duas abordagens que podem ser usadas para fornecer aos clientes URIs de recursos informacionais que descrevem recursos não-informacionais: URIs *hash* ou redirecionamentos HTTP 303 (BIEZER *et al.*, 2007).

Em URIs *hash*, por exemplo *http://example.com/people.rdf#alice*, a parte depois do *hash* é retirada e o URI resultante é desreferenciado. Um navegador de Dados Ligados irá receber a resposta (o arquivo RDF, neste caso), e encontrar triplas que dizem mais sobre o recurso não-informacional. A desvantagem desta abordagem é que os URIs gerados não são bons nomes, pois há uma referência a um formato de representação específica. Ao renomear o arquivo RDF mais tarde ou dividir os dados em vários arquivos, todos os identificadores vão mudar e os *links* existentes serão quebrados (BIEZER *et al.*, 2007).

Em redirecionamentos HTTP 303, em vez de enviar uma representação do recurso, o servidor envia ao cliente o URI de um recurso informacional que descreve o recurso não-informacional, utilizando o código de resposta HTTP 303 *See Other*. Em uma segunda etapa, o cliente desreferencia este novo URI e obtém uma representação que descreve o recurso não-informacional original (BIEZER *et al.*, 2007).

Redirecionamentos HTTP 303 podem ser usados juntamente com Negociação de Conteúdo, apresentada na seção seguinte.

3.4.8. Negociação de Conteúdo

Navegadores HTML geralmente exibem representações RDF como código bruto, ou simplesmente realizam o *download* dos arquivos sem exibi-los. Fornecer uma representação HTML apropriada da representação RDF de um recurso ajuda os usuários a descobrirem a que se refere um URI (BIEZER *et al.*, 2007).

Isto pode ser conseguido usando um mecanismo HTTP chamado Negociação de Conteúdo[34], "o processo de selecionar a melhor representação de uma dada resposta quando há múltiplas representações disponíveis" (FIELDING *et al.*, 1999).

Clientes HTTP enviam cabeçalhos juntamente com cada solicitação para indicar que tipo de representação eles preferem. Servidores podem inspecionar os cabeçalhos e selecionar uma resposta adequada. Se os cabeçalhos indicam que o cliente prefere HTML, o servidor pode gerar uma representação em HTML. Se o cliente prefere RDF, o servidor pode gerar RDF (HEATH *et al.*, 2008).

Negociação de Conteúdo para os recursos não-informacionais é geralmente implementado da seguinte forma: quando um URI que identifica um recurso não-

[34] http://www.w3.org/Protocols/rfc2616/rfc2616-sec12.html

informacional é desreferenciado, o servidor envia um redirecionamento 303 para um recurso informacional adequado para o cliente (BIEZER *et al.*, 2007).

Para ilustrar melhor, a figura 3.6 mostra a Negociação de Conteúdo para um cliente solicita uma representação RDF.

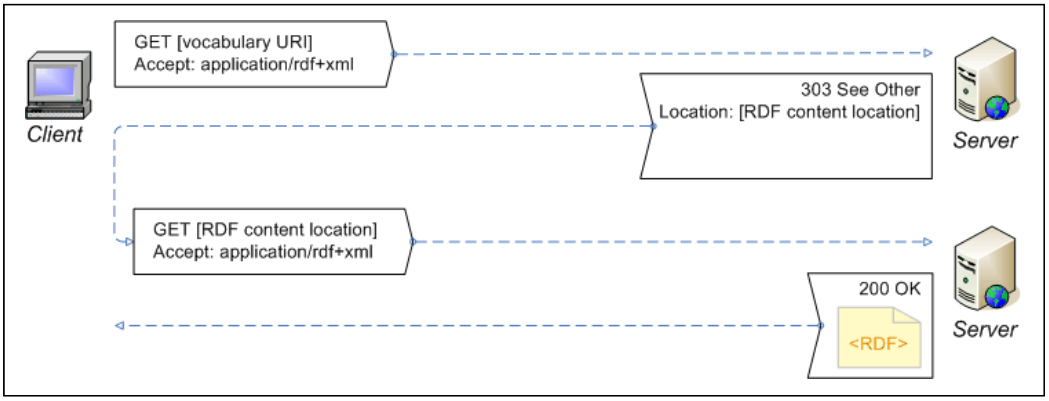


Figura 3.6: Exemplo de negociação de conteúdo (BIEZER *et al.*, 2007)

3.5. Linking Open Data

O exemplo mais visível da adoção e aplicação dos princípios de Dados Ligados é o projeto *Linking Open Data*[35], um esforço aberto e colaborativo apoiado pelo grupo W3C SWEO (*Semantic Web Education and Outreach Group*) [36] (BIZER *et al.*, 2008).

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 (BIZER *et al.*, 2008).

O projeto começou no início de 2007 com um número relativamente pequeno de *data sets* e de participantes, e desde então vem crescendo muito em termos de tamanho, impacto e colaboradores (HAUSENBLAS, 2009b). Este crescimento foi possível graças à sua natureza aberta, onde qualquer pessoa pode publicar *data sets* de acordo com os princípios de Dados Ligados e interligá-los com *data sets* existentes (BIZER *et al.*, 2009).

Uma indicação do tamanho do projeto é fornecida na figura 3.2, que mostra a nuvem de dados ligados gerada pelos diferentes *data sets* do projeto. Cada nó do diagrama representa um *data set* distinto, e cada arco indica que *links* existem entre itens nos *data sets* conectados.

[35] http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData
[36] http://www.w3.org/2001/sw/sweo/)

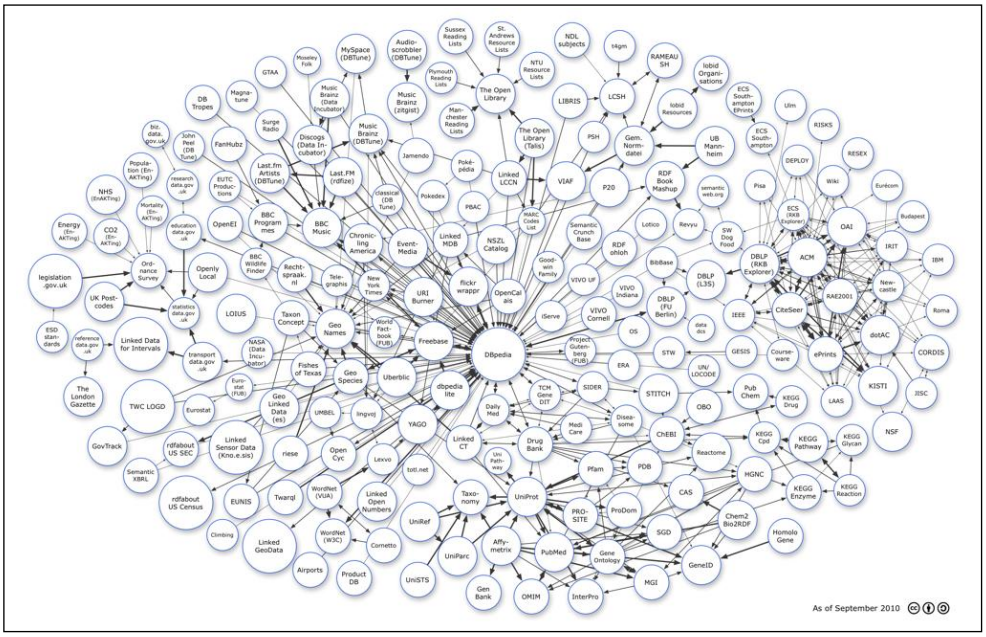


Figura 3.2: Diagrama de nuvens do projeto LOD (CYGANIAK e JENTZSCH, 2010)

Segundo estatísticas encontradas na página inicial do projeto, a Web de Dados atualmente consiste em 203 *data sets*, com mais de 25 bilhões de triplas RDF, interligadas por aproximadamente 395 milhões de *links* RDF (Setembro de 2010).

O conteúdo da nuvem é de natureza diversa, incluindo dados sobre localizações geográficas, pessoas, empresas, governos, livros, publicações científicas, filmes, programas de rádio e televisão, música, dados estatísticos, críticas, entre outros (BIZER *et al.*, 2009).

Exemplos de *data sets* de dados governamentais são: *Data.gov*, com mais de 1000 data sets do governo dos Estados Unidos disponíveis como Dados Ligados; *Data.gov.uk*, com mais de 3600 data sets do governo do Reino Unido disponíveis como Dados Ligados; *Data.gov Wiki*[37], com mais de 5 bilhões de triplas convertidas de *data sets* publicados no portal *Data.gov*; *Watchdog.net*[38], que coleta dados de políticos dos Estados Unidos; e o *GovTrack.us*[39], que publica Dados Ligados sobre membros do congresso dos Estados Unidos, bem como contas, comissões e votos.

Alguns *data sets* provêm URIs e descrições RDF para muitas entidades e conceitos comuns, sendo conectados a um grande número de outros *data sets*. Por exemplo, o *DBpedia*[40] consiste de triplas RDF extraídas de artigos da *Wikipedia*,

37 http://data-gov.tw.rpi.edu/wiki
38 http://watchdog.net/
39 http://www.govtrack.us/
40 http://dbpedia.org/

enquanto o *Geonames*[41] provê descrições RDF de diversas localizações geográficas do mundo todo (BIZER *et al.*, 2009).

3.6. Aplicações

Com um volume significativo de Dados Ligados sendo publicados na Web, inúmeras pesquisas e esforços estão sendo realizados para construir aplicações que exploram esta *Web* de Dados (BIZER *et al.*, 2009), apresentadas nas seções seguintes.

3.6.1. Aplicações Específicas

Muitos serviços estão sendo desenvolvidos oferecendo uma funcionalidade mais específica através do *mashup* de dados de diferentes *data sets* (BIZER *et al.*, 2009).

A *BBC Programmes*[42] e a *BBC Music*[43] estão usando tecnologias de Dados Ligados juntamente com o *DBPedia* e o *MusicBrainz* para conectar entidades equivalentes e aumentar o conteúdo com dados adicionais (HAUSENBLAS, 2009b).

O *Revyu*[44] é um site de críticas e avaliações baseado nos princípios de Dados Ligados. Além de publicar Dados Ligados, utiliza dados de outros *data sets* para melhorar a experiência dos seus usuários. Por exemplo, quando uma correspondência de algum filme é encontrada no *DBPedia*, informações adicionais são extraídas (HAUSENBLAS, 2009b).

O *DBpedia Mobile* é uma aplicação para *iPhone* e outros aparelhos móveis. Baseado na posição GPS (*Global Positioning System*) atual, mostra os locais próximos com informações do *DBPedia*, críticas associadas do *Reyvu* e fotos relacionadas através de um *wrapper* de Dados Ligados sobre a API de compartilhamento de fotos *Flickr*[45] (HAUSENBLAS, 2009b).

3.6.2. Motores de Busca e Indexadores

Há motores de busca de Dados Ligados que navegam a *Web* de dados, seguindo *links* entre fontes de dados e fornecendo recursos de consulta sofisticados sobre os dados, semelhantes aos utilizados pelos bancos de dados relacionais convencionais (BIZER *et al.*, 2009).

41 http://www.geonames.org/ontology/
42 http://www.bbc.co.uk/programmes
43 http://www.bbc.co.uk/music
44 http://revyu.com
45 http://www.flickr.com

Os motores de busca como o *Falcons*[46] e o *SWSE*[47] permitem serviços de busca orientados para os usuários humanos, fornecendo um resumo da entidade que o usuário seleciona a partir da lista de resultados, juntamente com os dados complementares estruturados extraídos da *Web* e *links* para entidades relacionadas (BIZER *et al.*, 2009).

Mecanismos como o *Swoogle*[48], *Sindice*[49] e *Watson*[50] fornecem APIs através das quais aplicações de Dados Ligados podem descobrir documentos RDF na *Web* que fazem referência a um certo URI ou contêm certa palavra-chave (BIZER *et al.*, 2009).

3.6.3. Navegadores

Assim como os navegadores Web tradicionais permitem aos usuários navegarem entre as páginas HTML seguindo os *links* de hipertexto, navegadores de Dados Ligados permitem navegar entre fontes de dados seguindo *links* RDF. Isso permite ao usuário começar com uma fonte de dados e, em seguida, passar por potencialmente inesgotáveis fontes de dados conectadas por *links* RDF (BIZER *et al.*, 2008).

Alguns exemplos de navegadores de Dados Ligados são: *The Tabulator*[51], *Disco Hyperdata Browser*[52], *OpenLink Data Web Browser*[53] e *Zitgist Data Viewer*[54].

3.7. Publicação

Existem várias maneiras de publicar Dados Ligados na *Web* dependendo de vários fatores, tais como a do tipo da informação, a quantidade de dados, a forma de armazenamento e a quantidade de vezes que os dados mudam (BIZER *et al.*, 2007).

Várias ferramentas de publicação foram desenvolvida para ajudar os editores a lidarem com detalhes técnicos e garantir que os dados sejam publicados de acordo com as práticas de Dados Ligados (BIZER *et al.*, 2009) .

A maneira mais simples de publicar Dados Ligados é produzir arquivos RDF estáticos e enviá-los para um servidor *Web*. Essa abordagem é geralmente escolhida em situações onde os arquivos RDF são criados manualmente ou são geradas por programas que apenas produzem saídas para arquivos (BERNERS-LEE, 2008).

---

46 http://ws.nju.edu.cn/falcons/
47 http://swse.org/
48 http://swoogle.umbc.edu/
49 http://www.sindice.com/
50 http://watson.kmi.open.ac.uk/WatsonWUI/
51 http://www.w3.org/2005/ajar/tab
52 http://www4.wiwiss.fu-berlin.de/bizer/ng4j/disco/
53 http://demo.openlinksw.com/DAV/JS/rdfbrowser/index.html
54 http://dataviewer.zitgist.com/

No caso de bancos de dados relacionais, há uma série de ferramentas livres para publicá-los como Dados Ligados, como por exemplo o *D2RServer*[55] (BERNERS-LEE, 2009). O servidor D2R realiza um mapeamento do banco de dados e gera uma representação de Dados Ligados, fornecendo também um *endpoint* para a realização de consultas SPARQL. Alternativamente, as ferramentas *Triplify*[56] e *OpenLink Virtuoso*[57] podem ser usadas (BIZER *et al.*, 2007).

Se a informação é representada como CSV, planilhas eletrônicas ou BibTEX[58], os dados podem ser convertidos para RDF usando uma ferramenta *RDFizing*. Depois disso, os dados podem ser armazenados em um repositório RDF. Como muitos repositórios RDF ainda não implementam interfaces de Dados Ligados, um repositório que fornece um *endpoint* SPARQL pode ser utilizado juntamente com a ferramenta *Pubby*[59], que permite adicionar uma interface de Dados Ligados sobre o *endpoint* SPARQL (BIZER *et al.*, 2007).

Se os dados estiverem em XML, um script pode ser criado para converter cada arquivo XML em RDF, usando por exemplo XSLT[60] ou outra linguagem (BERNERS-LEE, 2009).

Se os dados estiverem em um formato proprietário, um script pode ser criado para obter os dados e convertê-los para uma das formas padrão de Dados Ligados (BERNERS-LEE, 2009).

No caso de dados disponível na *Web* através de APIs, *wrappers* de Dados Ligados podem ser implementados da seguinte forma: URIs HTTP devem ser atribuídas para os recursos não-informacionais sobre os quais a API fornece dados. Quando um desses URIs é desreferenciado, o *wrapper* reescreve a solicitação do cliente, os resultados são transformados em RDF e enviados de volta para o cliente (BIZER *et al.*, 2007).

No caso de *websites*, os scripts podem ser construídas ou alterados usando RDFa para que os dados que estão por trás de cada página possam ser extraídos por outros como dados do modelo RDF. Outra alternativa é construir para cada página *Web* uma página paralela com os dados no modelo RDF, adicionando os detalhes técnicos de Dados Ligados (LEE, 2009; BIZER *et al.*, 2007; HEATH *et al.*, 2008).

---

55 http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/
56 http://triplify.org/
57 http://virtuoso.openlinksw.com/
58 http://www.bibtex.org/
59 http://www4.wiwiss.fu-berlin.de/pubby/
60 http://www.w3.org/TR/xslt20/

As seções seguintes descrevem diversas práticas para a publicação de informações como Dados Ligados na *Web*.

3.7.1. Escolha de URIs

Ao publicar os Dados Ligados, bons URIs devem ser escolhidos para os recursos. Eles devem ser bons nomes para que outros possam usá-los de forma segura. Além disso, deve ser considerada a infra-estrutura técnica para torná-los desreferenciáveis (BIZER et al., 2007) .

Segundo Bizer *et al.* (2007) e Heath (2009), deve-se ter em mente:

- Usar URIs HTTP para tudo. O esquema "http://" é o único amplamente suportado nas ferramentas e infra-estruturas atuais;
- Definir os URIs em um *namespace* HTTP próprio, e não no *namespace* de outros;
- Abstrair detalhes de implementação. Nomes pequenos e simples são melhores;
- Tentar manter os URIs estáveis e persistentes. A mudança posterior de URIs quebrará quaisquer *links* pré-estabelecidos;
- Os URIs escolhidos são limitadas pelo ambiente técnico;
- Quando for preciso usar chaves-primárias para garantir que cada URI é único, usar uma chave que é significativa dentro do seu domínio sempre que possível.

Geralmente são gerados três URIs relacionados a um único recurso não-informacional: um identificador para o recurso; um identificador para o recurso informacional adequado para navegadores HTML; e um identificador para o recurso informacional adequado para navegadores RDF (BIZER *et al.*, 2007).

A figura 3.3 mostra alguns exemplos de bons URIs que podem ser utilizados ao publicar Dados Ligados.

```
http://dbpedia.org/resource/Berlin <- recurso não-informacional
http://dbpedia.org/page/Berlin     <- página HTML
http://dbpedia.org/data/Berlin     <- dados RDF

http://dbpedia.org/Berlin          <- recurso não-informacional
http://dbpedia.org/Berlin.html     <- página HTML
http://dbpedia.org/Berlin.rdf      <- dados RDF

http://dbpedia.org/Berlin          <- recurso não-informacional
http://dbpedia.org/Berlin/html     <- página HTML
http://dbpedia.org/Berlin/rdf      <- dados RDF
```

Figura 3.3: Exemplos de bons URIs

3.7.2. Escolha de Vocabulários

Usar RDF implica a utilização de ontologias: decidir quais recursos são importantes dentro da aplicação e quais são as suas propriedades (TENNISON e SHERIDAN, 2008).

Para que aplicações clientes processem os dados mais facilmente, é considerado 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 para cobrir todas as classes e propriedades não podem ser encontrados em vocabulários existentes (HAUSENBLAS, 2009a; BIZER *et al.*, 2007; BERNERS-LEE, 2009).

Existe um conjunto de vocábulos conhecidos na comunidade da Web Semântica, como o *Friend-of-a-Friend*[61] (FOAF), para descrever pessoas; *Dublin Core*[62] (DC), para definir atributos de metadados em geral; *GeoNames* (GEO) para descrer dados sobre localizações geográficas; *Description of a Project*[63] (DOAP), para descrever projetos e *Creative Commons*[64] (CC), para descrever termos de licença (BIZER *et al.*, 2007).

Para ajudar a encontrar vocabulários, indexadores semânticos ou serviços dedicados podem ser usados, como o *Talis Schemacache*[65], *SchemaWeb.info*[66], *Falcons Concept Search*[67] ou *OntoSelect*[68] (HAUSENBLAS, 2009b; HEATH *et al.*, 2008).

[61] http://www.foaf-project.org/
[62] http://dublincore.org/
[63] http://*usefulinc.com/ns/doap*
[64] http://creativecommons.org/ns
[65] http://schemacache.test.talis.com/
[66] http://www.schemaweb.info/

Quando novos termos tiverem que ser definidos, deve ser publicado juntamente um arquivo usando as linguagens RDFS ou OWL. Ferramentas como a *Protégé*[69], *Neologism*[70] e *OpenVocab*[71] podem ser usadas para ajudar nesse processo (BIZER *et al.*, 2007; HEATH *et al.*, 2008) .

Sempre que possível, complemente vocabulários existentes com termos adicionais (no seu próprio namespace) para representar os seus dados conforme necessário (BIZER *et al.*, 2007). Fornecer mapeamentos para outros termos ajuda a promover o nível de intercâmbio na *Web* de Dados. Propriedades comuns para isso são: *rdfs:subClassOf*, *rdfs:subPropertyOf*, *owl:equivalentClass*, *owl:equivalentProperty* e *owl:inverseOf* (BIZER *et al.*, 2007; HEATH *et al.*, 2008).

É essencial que os URIs sejam desreferenciáveis para que os clientes possam olhar a definição de um termo (BIZER *et al.*, 2007).

Informações importantes devem ser fornecidas tanto para humanos quanto para máquinas adicionando propriedades como *rdfs:comments* e *rdfs:label* para cada termo inventado (BIZER *et al.*, 2007).

A listagem 3.6 mostra um exemplo que contem a definição de uma classe e uma propriedade seguindo as regras citadas acima.

```
# Definition of the class "Lover"
<http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LoveVocabulary#Lover>
    rdf:type rdfs:Class ;
    rdfs:label "Lover"@en ;
    rdfs:comment "A person who loves somebody."@en ;
    rdfs:subClassOf foaf:Person .
```

Listagem 3.6: Definições de novas classes e propriedades (BIZER *et al.*, 2007)

A *Web* de dados depende, portanto, de uma abordagem de integração de dados baseada em uma mistura da utilização de vocabulários comuns juntamente com termos de fonte de dados específicos, conectados por mapeamentos conforme necessário (BIZER *et al.*, 2009).

67 http://ws.nju.edu.cn/falcons/conceptsearch/index.jsp
68 http://sioc-project.org/node/192
69 http://protege.stanford.edu/
70 http://neologism.deri.ie/
71 http://open.vocab.org/

Estas são melhores práticas para a determinação vocabulários, mas não se deve esperar até que se tenha um esquema completo ou uma ontologia para publicar dados (BERNERS-LEE, 2009).

3.7.3. Adição de Metadados

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 (HARTIG, 2009 *apud* BIZER *et al.*, 2008). Para que os clientes usem os dados em termos legais claros, cada documento RDF deve conter a licensa sob a qual o conteúdo pode ser utilizado (BIZER *et al.*, 2007).

Os metadados podem ser fornecidos usando termos de vocabulários como o DC e FOAF, através das propriedades *dc:date*, *dc:publisher*, *dc:license*, *foaf:primaryTopic* e *foaf:topic*, conforme ilustrado na listagem 3.7.

```
# Metadata and Licensing Information
<http://dbpedia.org/data/Alec_Empire>
    rdfs:label "RDF description of Alec Empire" ;
    rdf:type foaf:Document ;
    dc:publisher <http://dbpedia.org/resource/DBpedia> ;
    dc:date "2007-07-13"^^xsd:date ;
    dc:rights
        <http://en.wikipedia.org/wiki/WP:GFDL> .
```

Listagem 3.7: Exemplo de metadados (BIZER *et al*., 2007)

3.7.4. Geração de Links

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).

Em um ambiente aberto como a *Web*, muitas vezes diferentes URIs identificam o mesmo recurso não-informacional. Esses URIs são chamados de URI *Aliases*. Para identificar as diferentes informações sobre um mesmo recurso, é uma prática comum o uso da propriedade *owl: sameAs* (BIEZER *et al.*, 2007). Outros predicados populares para a geração de links RDF são: *foaf:knows*, *foaf:homepage*, *foaf:topic*, *foaf:based_near*, *foaf:topic_interest*, *foaf:maker*, *foaf:made*, *foaf:depiction*, *foaf:page*,

foaf:primaryTopic, *geo:lat*, *rdfs:seeAlso* e *dc:isPartOf*. O domínio da aplicação irá determinar quais propriedades serão usadas como predicados (HEATH, 2009; HAUSENBLAS, 2009a; BIZER *et al.*, 2007).

A listagem 3.8 mostra alguns exemplos da utilização dessas propriedades para a realização de *links* RDF.

```
<http://www.w3.org/People/Berners-Lee/card#i>
    owl:sameAs <http://dbpedia.org/resource/Tim_Berners-Lee>;
    foaf:knows <http://www.w3.org/People/Connolly/#me>;
    foaf:topic_interest <http://dbpedia.org/resource/Semantic_Web>.
```

Listagem 3.8: Exemplos de *links* RDF externos (BIZER *et al*., 2007).

Links RDF podem ser definidos manualmente, mas como as fontes de dados muitas vezes fornecem informações sobre grande número de entidades, é prática comum o uso de abordagens automatizadas ou semi-automatizada para gerar *links* RDF (BIZER *et al.*, 2007).

Para gerar *links* manualmente, basta identificar os *data sets* adequados para a ligação e procurar os URIs desejados. Serviços como *Uriqr*[72] ou *Sindice*[73] também podem ser utilizados para pesquisar URIs existentes sobre uma dada entidade (BIZER *et al.*, 2007).

Silk[74] e *LIMES*[75] são exemplos de ferramentas para gerar *links* RDF automaticamente. Porém, existe ainda uma falta de boas ferramentas e fáceis de usar para esse fim. Por isso, é prática comum implementar algoritmos de ligação de *data sets* específicos para gerar *links* RDF (BIZER *et al.*, 2007).

Em vários domínios, existem geralmente esquemas de nomeação aceitos para identificar unicamente as entidades. Por exemplo, no domínio de livros existe o número ISBN (*International Standard Book Number*). Se o *link* de origem e o *link* do *data set al.*vo ambos possuirem essa identificação, a relação entre as duas entidades pode ser facilmente realizada usando um algorítmo simples (BIZER *et al.*, 2009).

No caso onde não existem identificadores comuns entre *data sets*, *links* RDF são geradas com base na similaridade de entidades de ambos os conjuntos de dados. É

[72] http://uriqr.com/
[73] http://www.sindice.com/
[74] http://www4.wiwiss.fu-berlin.de/bizer/silk/
[75] http://aksw.org/Projects/limes

necessário empregar algoritmos mais complexos de ligação baseados nas propriedades dos recursos (BIZER *et al.*, 2009).

3.7.5. Mecanismos de Descoberta

A maneira padrão de descobrir Dados Ligados na *Web* é seguindo *links* RDF através dos dados que o cliente conhece. A fim de facilitar ainda mais a descoberta, mecanismos adicionais podem ser utilizados (BIZER *et al.*, 2007).

Para tornar mais fácil não só para as máquinas mas também para os usuários descobrirem os dados, deve-se adicionar o *data set* na Lista de *Data Sets* na Wiki ESW[76], incluindo exemplos de URIs de recursos interessantes (BIZER *et al.*, 2007).

Pode-se utilizar uma extensão *sitemap* para indicar onde o RDF está localizado e que meios alternativos são fornecidos para acessá-lo, como Dados Ligados, *endpoints* SPARQL e RDF *dump*s, conforme indicado na listagem 3.9. Clientes e *crawlers* podem usar essas informações para acessar dados RDF de maneira mais eficiente (HEATH *et al.*, 2008).

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset>
    <sc:dataset>
        <sc:datasetLabel> The Wiskii.com dataset </sc:datasetLabel>
        <sc:linkedDataPrefix> http://wiskii.com/ </sc:linkedDataPrefix>
        <sc:dataDumpLocation>
            http://wiskii.com/dump.nt.gz
        </sc:dataDumpLocation>
        <sc:sparqlEndpointLocation>
            http://wiskii.com/sparql
        </sc:sparqlEndpointLocation>
        <changefreq>daily</changefreq>
    </sc:dataset>
</urlset>
```

Listagem 3.9: Exemplo de *Sitemap* (HEATH *et al*., 2008)

Também faz sentido em alguns casos definir *links* de páginas existentes para dados RDF, por exemplo, de páginas pessoais no perfil FOAF. Essas ligações podem ser definidas usando o elemento HTML <*link*> no cabeçalho da página. Esses elementos são usados por extensões do navegador, como *Piggybank*[77] e *Semantic Radar*[78], para descobrir dados RDF na *Web* (BIZER *et al.*, 2007).

[76] http://esw.w3.org/TaskForces/CommunityProjects/LinkingOpenData/DataSets
[77] http://simile.mit.edu/wiki/Piggy_Bank
[78] http://sioc-project.org/firefox

A divulgação dos dados também pode ser melhorada ao registrar URIs com *Ping The Semantic Web*[79], um serviço de registro de documentos RDF na *Web*, que é utilizado por vários outros serviços e aplicações cliente (HEATH *et al.*, 2008).

Deve-se garantir que há *links* RDF externos apontando para URIs do *data set*, de modo que navegador e *crawlers* RDF possam encontrar os dados. Além disso, deve-se convencer proprietários de fontes de dados relacionados à auto-gerar ou utilizar *links* RDF para URIs do *data set* (BIZER *et al.*, 2007).

3.7.6. Realização de Testes

Depois de publicar informações como Dados Ligados, deve-se testar se elas podem ser acessadas corretamente (HEATH *et al.*, 2008).

Uma maneira fácil de testar é colocar vários URIs do *data set* no serviço de validação *Vapour Linked*[80], que gera um relatório detalhado sobre como os URIs reagem a diferentes solicitações HTTP (HEATH *et al.*, 2008).

Além disso, também é importante ver se a informações são mostradas corretamente em diferentes navegadores de Dados Ligados e se eles podem seguir *links* RDF através dos dados (BIZER *et al.*, 2007).

Se ocorrerem problemas, deve-se realizar testes com cURL[81] para garantir que ao desreferenciar os URIs são geradas respostas HTTP corretas. (BIEZER *et al.*, 2007).

Deve-se utilizar também o Serviço de Validação RDF do W3C[82] para se certificar de que são fornecidos documentos RDF/XML válidos (HEATH *et al.*, 2008).

3.8. Desafios

Para utilizar a Web como um único banco de dados global vários desafios devem ser superados. (BIZER *et al.*, 2009)

Quando Dados Ligados surgiram, o foco principal da comunidade era publicar dados e encontrar boas práticas para isso. Agora, outras questões importantes estão sendo abordadas, tais como usabilidade, qualidade, escalabilidade, desempenho, e confiabilidade (ALEXANDER, 2009; HAUSENBLAS, 2009b).

79 http://pingthesemanticweb.com/
80 http://vapour.sourceforge.net/
81 http://curl.haxx.se/
82 http://www.w3.org/RDF/Validator/

A *Web* de dados legíveis por máquina cria desafios e oportunidades para as interfaces de usuário e paradigmas de interação. Os navegadores e motores de busca atuais ainda não são utilizáveis pelos usuários comuns da *Web*. São necessários: visões amigáveis sobre a ordenação dos dados e fusão de propriedades; recursos mais avançados de análise de dados; e declarações sobre a proveniência e a confiabilidade dos dados (HAUSENBLAS, 2009b).

Hoje, a maioria das aplicações de Dados Ligados exibem dados de diferentes fontes, mas pouco é feito para integrá-los de uma forma significativa. Para isso é preciso mapear os termos dos diferentes vocabulários, bem como realizar a fusão de dados sobre a mesma entidade, resolvendo conflitos de dados (BIZER *et al.*, 2009).

Aplicações que consomem dados da Web devem ser capazes de acessar as especificações explícitas dos termos de licença em que os dados podem ser reutilizados e republicados. São necessários vocabulários apropriados e melhores práticas para utilizar metadados de licenciamento (BIZER *et al.*, 2009).

Segundo Stankovic (2009), *data sets* que poderiam ser interligados ainda não o são. Além disso, uma grande quantidade de dados que seriam fontes úteis estão faltando na LOD *Cloud* atual. Em alguns casos os tipos de dados necessários existem, mas as descrições de metadados não são suficientes finas ou detalhes necessários estão faltando.

Outro problema é que o conteúdo das fontes de Dados Ligados sofrem alterações constantemente. Atualmente, links RDF entre fontes de dados são atualizados apenas esporadicamente, o que leva a links quebrados apontando para URIs que já não são mantidos (BIZER *et al.*, 2009).

3.9. Considerações Finais

Neste capítulo foram abordadas as principais questões relacionadas aos Dados Ligados, bem como as principais tecnologias e práticas necessárias para publicá-los. É importante seguir as práticas apontadas para garantir uma melhor utilização dos dados.

O capítulo seguinte aborda a implementação proposta, que visa usar os conceitos até aqui expostos para a publicação de Dados Governamentais Abertos de políticos brasileiros utilizando as práticas de Dados Ligados.

4. Projeto: Data set de Políticos Brasileiros

Este capítulo apresenta o projeto que foi realizado utilizando os conceitos de Dados Governamentais Abertos e Dados Ligados.

A seção 4.1 fornece uma visão geral do projeto e sua arquitetura. A seção 4.2 apresenta mais detalhadamente a implementação e cada módulo do projeto. A seção 4.3 traz uma avaliação do projeto no que diz respeito aos princípios apresentados nas seções anteriores. Por fim, a seção 4.4 traz as considerações finais do capítulo.

4.1. Visão Geral

O objetivo do projeto é fornecer um novo *data set* com informações de políticos brasileiros coletadas de diferentes fontes utilizando as práticas de Dados Ligados e Dados Governamentais abertas.

Foram utilizados para a coleta de dados o *site* do Tribunal Superior Eleitoral (TSE)[83], o *site* do Senado Federal[84], o Portal da Câmara dos Deputados[85], o *site* da ONG (Organização Não Governamental) Políticos Brasileiros[86], o *site* Ficha Limpa[87] e o projeto Excelências[88] do *site* Transparência Brasil[89].

Como na maioria destes *sites* as informações estavam disponíveis somente em formato HTML, foi necessária a criação e utilização de *Web Crawlers* para extrair os dados de uma forma metódica e automatizada. Em alguns poucos casos, os dados estavam disponíveis em formato CSV.

Os dados extraídos das diferentes fontes foram inseridos em um banco de dados relacional criando uma base de dados única.

Em seguida, foi criada uma representação RDF/XML dos dados de cada recurso, sendo este o padrão utilizado para publicar Dados Ligados na Web. Para isso, foram levados em consideração diversos fatores como a escolha de URIs, definição de vocabulários e adição de metadados. Ao mesmo tempo, foram realizados *links* RDF para diferentes fontes de dados, prática que caracteriza os Dados Ligados.

83 http://www.tse.gov.br
84 http://www.senado.gov.br/
85 http://www.camara.gov.br/
86 http://politicosbrasileiros.com.br/
87 http://www.fichalimpa.org.br/
88 http://www.excelencias.org.br/
89 http://www.transparencia.org.br/

Em seguida, foram utilizados diferentes mecanismos de descoberta de forma a divulgar o *data set* e aperfeiçoar a consulta por máquinas. Foram realizados também testes para garantir que o projeto seguia os princípios e as práticas de Dados Ligados e Dados Governamentais Abertos.

Seguindo as orientações de Dados Ligados e Dados Governamentais Abertos, foi criada uma representação HTML para a visualização e consulta dos dados. Além disso, foram fornecidos também os dados brutos para possíveis reutilizações. Para enriquecer a experiência dos usuários foram gerados também gráficos sobre alguns dados.

As informações são extraídas diretamente da base de dados relacional e representadas dinamicamente em HTML ou RDF de acordo com a requisição do cliente e com os dados de cada recurso.

A figura 4.1 apresenta a arquitetura geral do projeto, mostrando os diferentes módulos e as ligações entre eles.

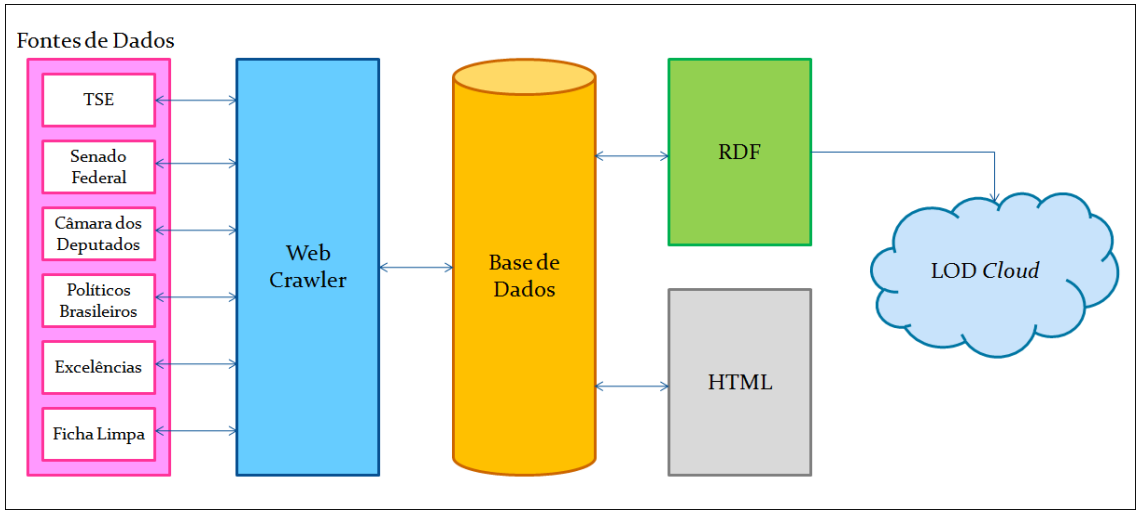


Figura 4.1: Arquitetura geral do projeto

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

Com o *data set* de dados de políticos brasileiros *online*, é possível consultar os dados, 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*.

4.2. Implementação

As seções seguintes apresentam mais detalhadamente cada módulo do projeto, bem como as práticas e recursos utilizados para cada fim.

4.2.1. Fontes de Dados

Em um primeiro momento, foi necessário pesquisar as fontes de dados de políticos brasileiros e os dados que seriam utilizados.

O *site* do TSE traz informações sobre todos os candidatos às eleições. Foram utilizados os dados pessoais e de divulgação de bens dos candidatos do ano de 2010.

O *site* do Senado Federal traz informações sobre os senadores em exercício e afastados, além de um histórico de senadores antigos. Foram extraídos os dados pessoais e parlamentares, lideranças, missões, pronunciamentos, comissões e proposições de candidatos em exercício e afastados.

O Portal da Câmara dos Deputados traz informações sobre os deputados atuais e antigos. Foram utilizados os dados pessoais e parlamentares, proposições e discursos de candidatos em exercício.

O *site* Ficha Limpa traz informações dos candidatos às eleições, e foi utilizado para complementar dados pessoais e fornecer dados de mandatos anteriores e afastamentos.

O *site* da ONG Políticos Brasileiros e o projeto Excelências do *site* Transparência Brasil foram utilizados para complementar os dados de políticos brasileiros em exercício.

4.2.2. Web Crawler

Apenas alguns dados pessoais dos políticos estão disponibilizados em formato aberto no *site* do TSE e no Portal da Câmara dos Deputados. Nos demais *sites* os dados estão disponíveis apenas para visualização.

Por esse motivo, foi necessária a criação e utilização de *Web Crawlers* para extrair os demais dados. A técnica para extrair dados *online* que não estão em formato aberto é denominada *screen scraping* ou raspagem de dados.

Existem diversos códigos e ferramentas que auxiliam nessa prática, em diferentes linguagens e plataformas. No projeto em questão, foi utilizada a linguagem

PHP (*Hypertext Preprocessor*) para a criação de *scripts* personalizados para extrair as diferentes informações de cada fonte de dados.

De maneira geral, os *scripts* funcionam da seguinte forma: são passados os URLs (*Uniform Resource Locators*) das páginas para uma função que pega todo o conteúdo e percorre os elementos do documento HTML através da linguagem XPath (*XML Path Language*), retornando o conteúdo dos elementos previamente programados no código através dos seus atributos e posicionamentos. Para percorrer todas as páginas foram realizados laços de repetição e utilizados os parâmetros que diferenciavam cada URL.

A listagem 4.1 mostra uma parte do código utilizado para realização da raspagem de dados de um dos elementos da declaração de bens dos candidatos do site do TSE.

```
function crawlDeclaracaoBens($url)
{
    $content=$this->getContent($url);
    $domain=$this->getDomain($url);
    $dom = new DOMDocument();
    @$dom->loadHTML($content);
    $xpath = new DOMXPath($dom);

    $elementos1 = $xpath->evaluate("//div[@id='div_bens']//td[2]");
    [...]
    for ($i = 0; $i < $elementos1->length - 1; $i++)
    {
        $elemento1 = $elementos1->item($i);
        $dados['descricao'][$i]=$elemento1->nodeValue;
    }
    [...]
    return  $dados;
}
```

Listagem 4.1: Parte do código do *Web Crawler* do projeto

Em alguns momentos, diferentes dados eram apresentados em um mesmo elemento HTML, sendo necessário utilizar mecanismos específicos para garantir a granularidade dos dados. Para isso, foram utilizadas funções específicas da linguagem PHP para separar os dados antes de inseri-los na base de dados. Outras dificuldades como a codificação dos caracteres e a formatação das datas também foram tratadas através de diferentes funções e conversões.

4.2.3. Base de Dados

Ao extrair os dados com o *Web Crawler* eles eram inseridos em uma base de dados relacional. Como cada fonte de dados utiliza uma abordagem diferente para armazenar e

apresentar os dados, foi necessário modelar a nova base de dados de maneira a conciliar essas abordagens sem perder ou modificar os dados, levando em consideração também a manutenibilidade da mesma. Para isso, foi preciso criar tabelas, campos e relacionamentos que representavam os dados das diferentes fontes de uma maneira única. O MySQL[90] foi utilizado como Sistema de Gerenciamento de Banco de Dados. A figura 4.2 apresenta a modelagem do banco de dados gerado.

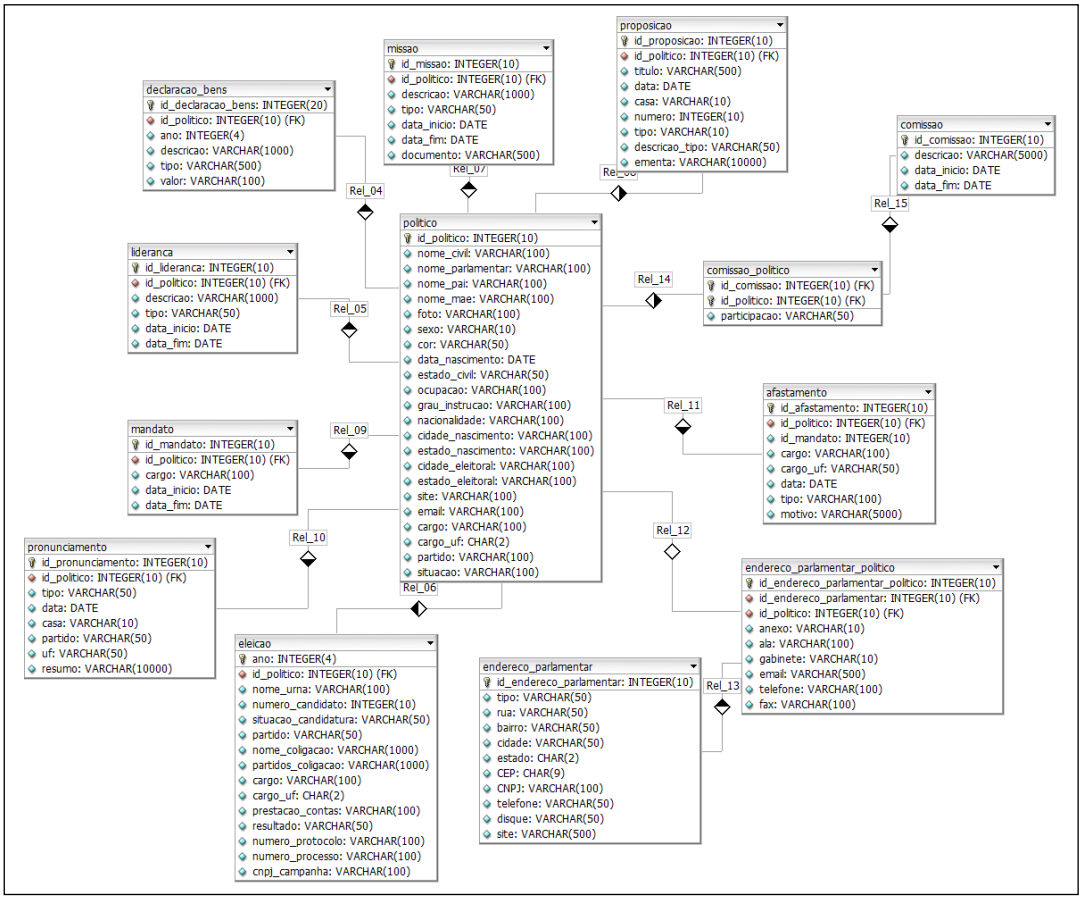


Figura 4.2: Modelagem do banco de dados do projeto

Não existia um identificador comum entre as diferentes fontes de dados, pois não é publicada junto com os dados dos políticos uma chave que garanta que eles são únicos, como o ISBN para os livros. Portanto, para identificar os dados que tratavam da mesma entidade foi necessário utilizar os diferentes dados dos políticos, como nome, partido e data de nascimento, de forma a garantir a consistência e evitar a duplicação. A falta de um identificador comum para os políticos acabou acarretando em uma limitação do projeto, pois não existe uma chave significativa dentro do domínio.

[90] http://www.mysql.com

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 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 tratados através de procedimentos de limpeza, integração e transformação.

A tabela 4.1 fornece uma visão mais ampla dos dados que foram coletados e dos tratamentos que tiveram que ser realizados em cada fonte de dados.

Fonte de Dados	Dados Coletados	Tratamento Realizado
TSE	Dados Pessoais (nome completo, data de nascimento, nacionalidade, naturalidade, grau de instrução, sexo, estado civil, site e ocupação), Declaração de Bens (descrição, tipo e valor) e Dados da Eleição (partido, coligação, cargo, limites de gastos da campanha, protocolo da campanha e CNPJ da campanha).	Dividir a naturalidade em cidade de nascimento e estado de nascimento para garantir a granularidade. Criar um campo de ano para a Declaração de Bens e para Dados da Eleição para garantir a manutenibilidade. Tratar a formatação das datas.
Senado Federal	Dados Pessoais (nome completo, data de nascimento, naturalidade), Dados Políticos (nome parlamentar, cargo, partido, estado), Endereço Parlamentar (ala, gabinete, rua, bairro, cidade, estado, CEP, CNPJ, email, telefone, fax, site), Lideranças (descrição, data início, data término), Missões (descrição, data início, data término, documento), Proposições (titulo, ementa, autor), Mandatos, Afastamento, Comissões (descrição, data início, data término, participação) e Pronunciamentos (tipo, Data, Casa, Partido, UF, Resumo).	Dividir os diversos dados de Lideranças, Missões, Proposições, Comissões, e Pronunciamentos para garantir a granularidade, além da naturalidade. Tratar a codificação dos caracteres e a formatação das datas. Criar um campo de tipo para as diversas tabelas para garantir a integração com outras fontes. Transformar as siglas dos partidos para o formato de representação padrão com espaçamento.
Câmara dos Deputados	Dados Pessoais (nome completo, data de nascimento, profissão), Dados Políticos (cargo, partido, estado), Endereço Parlamentar (anexo, gabinete, rua, bairro, cidade, estado, CEP, CNPJ, email, telefone, fax, site), Projetos (título, órgão, situação, data, ementa, explicação, despacho) e Discursos (data, sessão, fase, hora, sumário).	Dividir os diversos dados de Projetos e Discursos apresentados para garantir a granularidade. Tratar a codificação dos caracteres e a formatação das datas. Transformar as siglas dos partidos para o formato de representação padrão com espaçamento.

Políticos Brasileiros	Dados Pessoais (nome completo, sexo, data de nascimento, estado civil, nacionalidade, naturalidade, grau de instrução, ocupação) e Dados Políticos (cargo, partido, estado).	Dividir a naturalidade em cidade de nascimento e estado de nascimento para garantir a granularidade. Tratar a formatação das datas.
Ficha Limpa	Dados Pessoais (cor, site, cidade eleitoral, estado eleitoral), Mandatos (cargo, data início, data término) e Afastamentos (cargo, data, motivo).	Transformar os nomes dos estados para o formato de representação padrão através de siglas. Tratar a formatação das datas.
Excelências	Ocorrências (título, sigla, tipo, ano, número, descrição).	Dividir os diversos dados de Ocorrência apresentados para garantir a granularidade.

Tabela 4.1: Dados coletados e tratamentos realizados em cada fonte de dados

4.2.4. Representação RDF

Após obtermos todos os dados estruturados, a informação foi representada usando o modelo RDF e utilizando os princípios e práticas de Dados Ligados.

Primeiramente, foram resolvidos os detalhes técnicos, como a Negociação de Conteúdo e a desreferencia dos URIs, de forma a garantir que as representações dos recursos não-informacionais, no caso os políticos, fossem gerados corretamente através de bons URIs tanto para clientes RDF como para clientes HTML.

Para garantir a Negociação de Conteúdo, foi utilizado o *EasyPub*[91], um *script* em PHP para a publicação de RDFs disponibilizado no site do projeto Linked Data[92], sendo realizadas as devidas alterações no que diz respeito a estrutura do redirecionamento.

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. Exemplos de URIs utilizados são mostrados na Figura 4.3.

```
http://ligadonospoliticos.com.br/resource/1       <- recurso não-informacional
http://ligadonospoliticos.com.br/resource/1/html  <- página HTML
http://ligadonospoliticos.com.br/resource/1/rdf   <- dados RDF
```

Figura 4.3: Exemplo de URIs utilizados no projeto

[91] http://buzzword.org.uk/2009/easypub/
[92] http://linkeddata.org/

Portanto, quando o cliente entra com o URI que representa o recurso não-informacional, como por exemplo *http://ligadonospoliticos.com.br/resource/1*, é gerada uma representação de acordo com o cabeçalho enviado pelo cliente HTTP, sendo realizado o redirecionamento para os dados RDF ou para a página HTML.

Em seguida, foi realizada a escolha dos vocabulários que seriam utilizados para representar as propriedades dos recursos. Termos de vocabulários conhecidos como FOAF, BIO[93], PERSON[94], VCARD[95], DBPPROP[96], POL[97], BEING[98], TIME[99], BIBLIO[100] e DCTERMS[101] foram reutilizados sempre quando possível. Para encontrar os termos foi utilizado o serviço *Talis Schemacache* e realizadas pesquisas a dados RDF de outros *data sets*. Em alguns casos, diversos vocabulários definem o mesmo termo, sendo necessário utilizar parâmetros para escolher qual termo utilizar, como uma boa documentação e a utilização em outros *data sets*.

Não foram encontrados todos os termos necessários em outros vocabulários. Nesses casos, novos termos foram definidos representado por *polbr* sob o namespace "*http://ligadonospoliticos.com.br/politicobr/*". Foram publicados também arquivos RDFS descrevendo esses termos. A listagem 4.2 mostra um exemplo da definição de novas propriedades.

```
<rdf:RDF [...]>
    <rdf:Description
        rdf:about="http://ligadonospoliticos.com.br/politicobr/governmentalName">
        <rdfs:label xml:lang="en">governmentalName</rdfs:label>
        <rdfs:comment xml:lang="en">The governmental name of a person</rdfs:comment>
        <rdfs:isDefinedBy rdf:resource="http://ligadonospoliticos.com.br/politicobr"/>
    </rdf:Description>
</rdf:RDF>
```

Listagem 4.2: Exemplo de definições de novas propriedades para o projeto

Com os termos definidos, a informação foi representada no modelo RDF. 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

93 http://purl.org/vocab/bio/0.1/
94 http://models.okkam.org/ENS-core-vocabulary#
95 http://www.w3.org/2006/vcard/ns#
96 http://dbpedia.org/property/
97 http://www.rdfabout.com/rdf/schema/politico/
98 http://purl.org/ontomedia/ext/common/being#
99 http://pervasive.semanticweb.org/ont/2004/06/time#
100 http://purl.org/ontology/bibo/
101 http://purl.org/dc/terms/

pelos valores literais retirados da base de dados ou por URIs que de representam *links* RDF de outros recursos. Foram adicionados também metadados usando termos dos vocabulários DC e CC.

A listagem 4.3 mostra um exemplo de dados representados pelo modelo RDF gerado para um dado recurso. É possível também observar a reutilização de vocabulários, a utilização de termos criados e a adição de metadados.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:foaf="http://xmlns.com/foaf/0.1/"
         xmlns:bio="http://purl.org/vocab/bio/0.1/"
         xmlns:pol="http://www.rdfabout.com/rdf/schema/politico/"
         xmlns:dc="http://purl.org/dc/elements/1.1/"
         xmlns:polbr="http://ligadonospoliticos/politicobr/"
         [...] >
    <rdf:Description rdf:about="http://ligadonospoliticos.com.br/resource/1">
        [...]
        <dc:creator rdf:resource="http:// ligadonospoliticos.com/foaf.rdf"/>
        <foaf:name>DILMA VANA ROUSSEFF</foaf:name>
        <pol:party>PT</pol:party>
        <bio:mother>Dilma Jane Silva</bio:mother>
        <polbr:maritalstatus>Divorciado</polbr:maritalstatus>
        [...]
     </rdf:Description>
</rdf:RDF>
```

Listagem 4.3: Exemplo de dados representados pelo modelo RDF do projeto

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, são apresentadas como recursos para outras fontes de dados.

Dessa forma, *links* RDF foram gerados com os *data sets* DBPedia, GeoNames, Freebase[102], World Factbook[103], UMBEL (*Upper Mapping and Binding Exchange Layer*)[104] e YAGO[105] conforme ilustra a listagem 4.4.

[102] http://www.freebase.com/
[103] http://www4.wiwiss.fu-berlin.de/factbook/
[104] http://www.umbel.org/
[105] http://mpii.de/yago

```
<being:place-of-birth rdf:resource="http://dbpedia.org/resource/Belo_Horizonte"/>
<polbr:state-of-birth rdf:resource="http://www.geonames.org/3457153"/>
<person:occupation rdf:resource="http://rdf.freebase.com/rdf/en.economist"/>
<skos:subject rdf:resource="http://dbpedia.org/resource/Category:Living_people"/>
<rdf:type rdf:resource="http://dbpedia.org/class/yago/BrazilianPoliticians"/>
<vcard:country-name
       rdf:resouce="http://www4.wiwiss.fu-berlin.de/factbook/resource/Brazil"/>
<foaf:page rdf:resource="http://en.wikipedia.org/wiki/Dilma_Rousseff"/>
<owl:sameAs rdf:resource="http://dbpedia.org/page/Dilma_Rousseff"/>
```

Listagem 4.4: Exemplos de *links* RDF gerados no projeto

A figura 4.4 apresenta as ligações entre o *data set* Ligado nos Políticos e alguns *data sets* do projeto *Linking Open Data*.

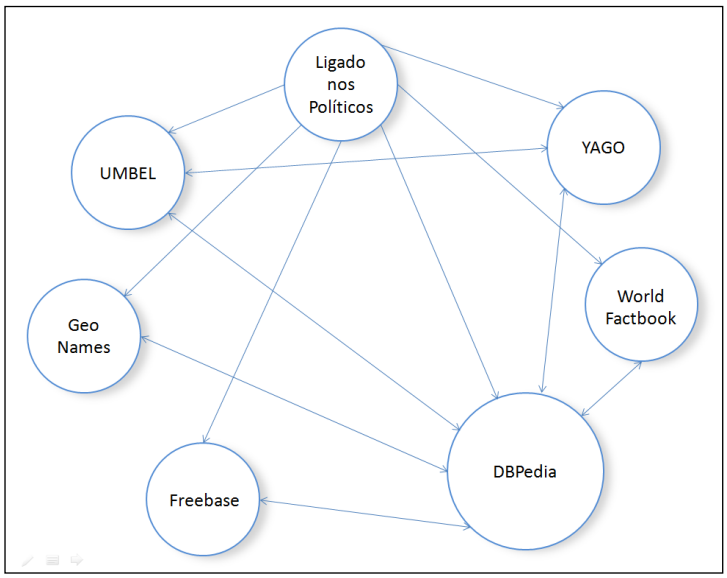


Figura 4.4: Ligações entre o *data set* Ligado nos Políticos e outros *data sets* do LOD

Como não existiam identificadores comuns entre os *dada sets*, foram utilizadas abordagens manuais e semi-automatizada para gerar os *links* RDF, esta última através de práticas semelhantes às utilizadas no *Web Crawler*, percorrendo as páginas das outras fontes de dados, comparando as propriedades dos recursos e extraindo os *links*.

As informações são retiradas da base de dados e inseridas dinamicamente no modelo de acordo com o recurso solicitado. Estruturas condicionais foram inseridas para garantir que nós em branco não fossem gerados.

A figura 4.5 mostra uma tela de exemplo da representação RDF gerada em um navegador comum.

Figura 4.5: Tela de exemplo da representação RDF gerada em um navegador comum

A fim de facilitar descoberta dos dados por máquinas e humanos, mecanismos adicionais foram utilizados. O *data set* foi adicionado na Lista de *data sets* na Wiki ESW, como apresentado na figura 4.6.

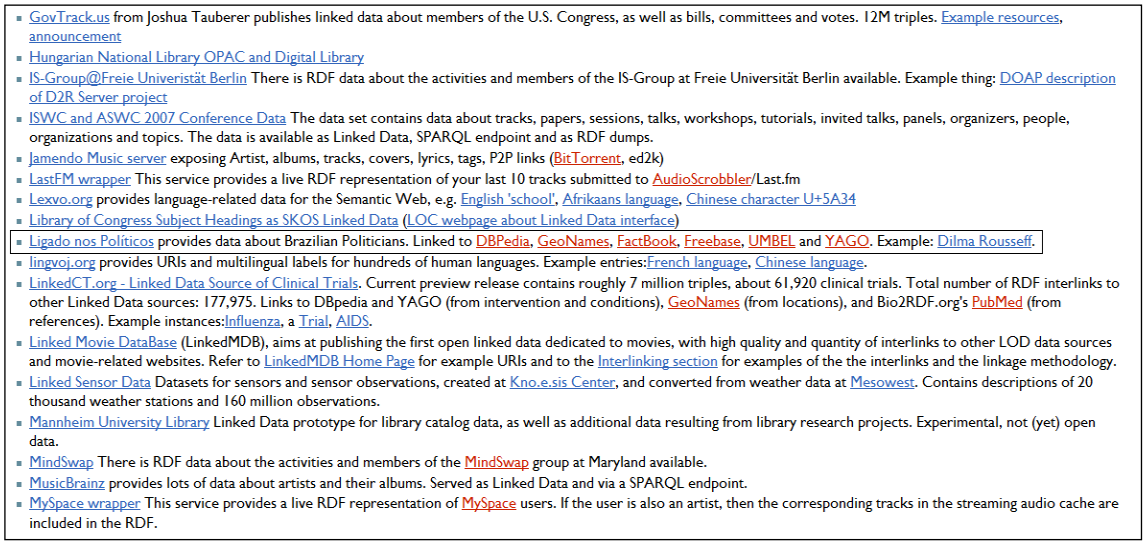


Figura 4.6: Informações do *data set* Ligado nos Políticos no site do LOD

Também foi realizado um mapa do site para indicar onde o RDF está localizado, conforme mostra a listagem 4.5.

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset>
    <sc:dataset>
        <sc:datasetLabel> Ligado nos Politicos </sc:datasetLabel>
        <sc:linkedDataPrefix>
            http://ligadonospoliticos.com.br/
        </sc:linkedDataPrefix>
        <changefreq>monthly</changefreq>
    </sc:dataset>
</urlset>
```

Listagem 4.5: *Sitemap* do projeto

Para testar se os dados podem ser acessadas corretamente e se eles podem seguir *links* RDF, foi utilizado o *The Tabulator Extension*[106], uma extensão para o navegador *Firefox*[107] que provê uma interface para Dados Ligados baseada no navegador *The Tabulator*.

A figura 4.7 fornece uma visão de um exemplo de tela gerada.

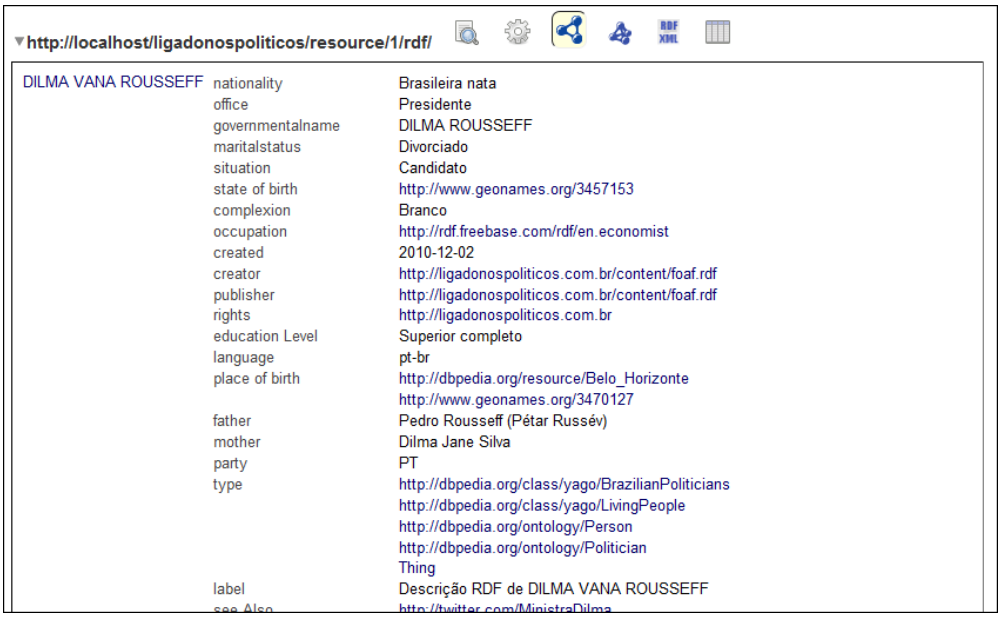
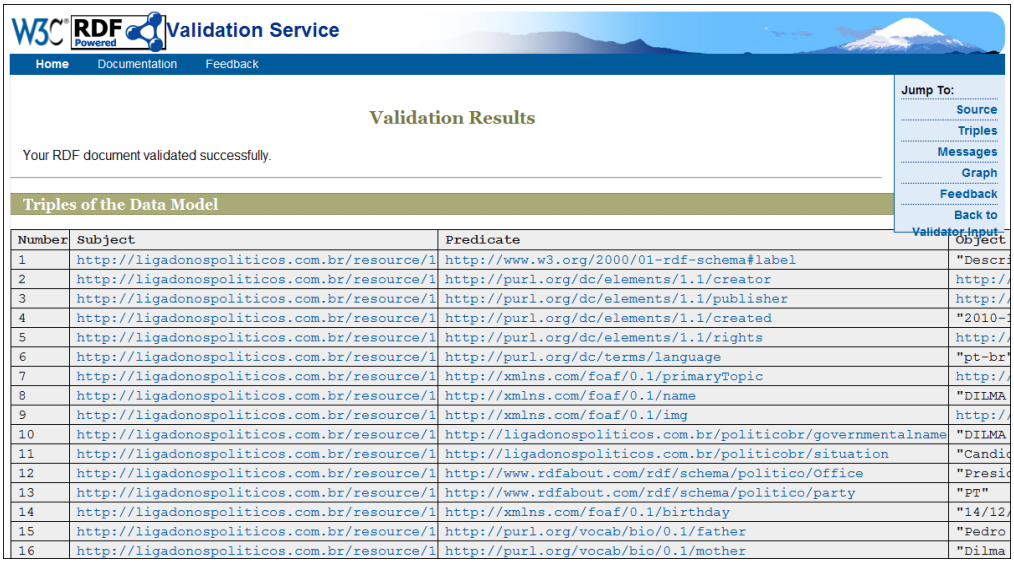


Figura 4.7: Tela de exemplo gerada pelo *The Tabulator Extension*

Também foi utilizado o serviço de validação RDF do W3C para certificar que são fornecidos documentos RDF/XML válidos, conforme ilustra a figura 4.8.

106 http://dig.csail.mit.edu/2007/tab/
107 http://www.mozilla.com/



W3C RDF Validation Service

Home Documentation Feedback

Jump To: Source Triples Messages Graph Feedback Back to Validator Input

Validation Results

Your RDF document validated successfully.

Triples of the Data Model

Number	Subject	Predicate	Object
1	http://ligadonospoliticos.com.br/resource/1	http://www.w3.org/2000/01-rdf-schema#label	"Descri
2	http://ligadonospoliticos.com.br/resource/1	http://purl.org/dc/elements/1.1/creator	http:/
3	http://ligadonospoliticos.com.br/resource/1	http://purl.org/dc/elements/1.1/publisher	http:/
4	http://ligadonospoliticos.com.br/resource/1	http://purl.org/dc/elements/1.1/created	"2010-1
5	http://ligadonospoliticos.com.br/resource/1	http://purl.org/dc/elements/1.1/rights	http:/
6	http://ligadonospoliticos.com.br/resource/1	http://purl.org/dc/terms/language	"pt-br"
7	http://ligadonospoliticos.com.br/resource/1	http://xmlns.com/foaf/0.1/primaryTopic	http:/
8	http://ligadonospoliticos.com.br/resource/1	http://xmlns.com/foaf/0.1/name	"DILMA
9	http://ligadonospoliticos.com.br/resource/1	http://xmlns.com/foaf/0.1/img	http:/
10	http://ligadonospoliticos.com.br/resource/1	http://ligadonospoliticos.com.br/politicobr/governmentalname	"DILMA
11	http://ligadonospoliticos.com.br/resource/1	http://ligadonospoliticos.com.br/politicobr/situation	"Candid
12	http://ligadonospoliticos.com.br/resource/1	http://www.rdfabout.com/rdf/schema/politico/Office	"Presid
13	http://ligadonospoliticos.com.br/resource/1	http://www.rdfabout.com/rdf/schema/politico/party	"PT"
14	http://ligadonospoliticos.com.br/resource/1	http://xmlns.com/foaf/0.1/birthday	"14/12/
15	http://ligadonospoliticos.com.br/resource/1	http://purl.org/vocab/bio/0.1/father	"Pedro
16	http://ligadonospoliticos.com.br/resource/1	http://purl.org/vocab/bio/0.1/mother	"Dilma

Figura 4.8: Tela de exemplo da utilização do serviço de validação RDF do W3C

4.2.5. Representação HTML

A representação HTML facilita a descoberta dos dados do *data set* e prove uma interface de consulta e visualização para os usuários. A página inicial do site é apresentada na figura 4.9.

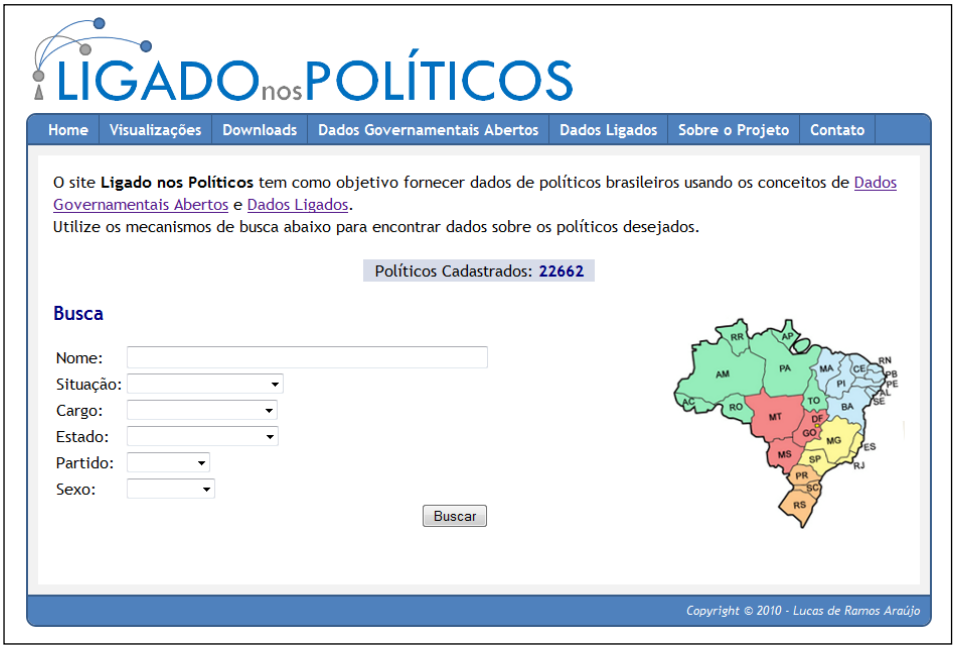


Figura 4.9: Tela da página inicial do site Ligado nos Políticos

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.

É fornecido também um mapa onde o usuário pode selecionar por Estado os políticos que deseja buscar.

Para facilitar a reutilização dos dados, eles também são fornecidos em seu formato bruto na página de *Downloads*.

São oferecidas também páginas com os principais conceitos de Dados Ligados e Dados Governamentais Abertos, além de uma página com uma descrição mais detalhada do projeto e uma página de Contato.

Após realizada a busca, caso haja registros na tabela de acordo com a consulta, são exibidos os políticos encontrados, identificados por nome, situação, cargo, Estado e partido, conforme indicado na figura 4.10. Caso contrário é exibida uma mensagem junto com o mecanismo de busca onde o usuário pode refazer a pesquisa.

Figura 4.10: Tela de exemplo do resultado da busca do site Ligado nos Políticos

É gerado um *link* para cada recurso não-informacional com o seu respectivo URI. Ao selecionar o político, é exibida a representação adequada de acordo com o cabeçalho HTTP enviado pelo cliente. Navegadores de Dados Ligados exibem a representação RDF enquanto navegadores comuns exibem a representação HTML, esta última apresentada na figura 4.11.

Figura 4.11: Tela de exemplo da representação HTML no site Ligado nos Políticos

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.

Para melhorar a experiência dos usuários e conferir novos valores aos dados, foram gerados gráficos utilizando os *site Many Eyes*[108] da IBM (*International Business Machines*)[109], que permite aos usuários realizarem o envio de dados e produzirem diferentes representações gráficas. Os dados enviados e as representações ficam abertas para a reutilização e visualização.

Para isso, foram realizadas primeiramente diferentes visões na base de dados para gerar os resultados necessários, como por exemplo, a quantidade de proposições de políticos cadastrados, e em seguida utilizada uma das visualizações disponíveis, conforme mostra a figura 4.12.

[108] http://www-958.ibm.com/
[109] http://www.ibm.com/

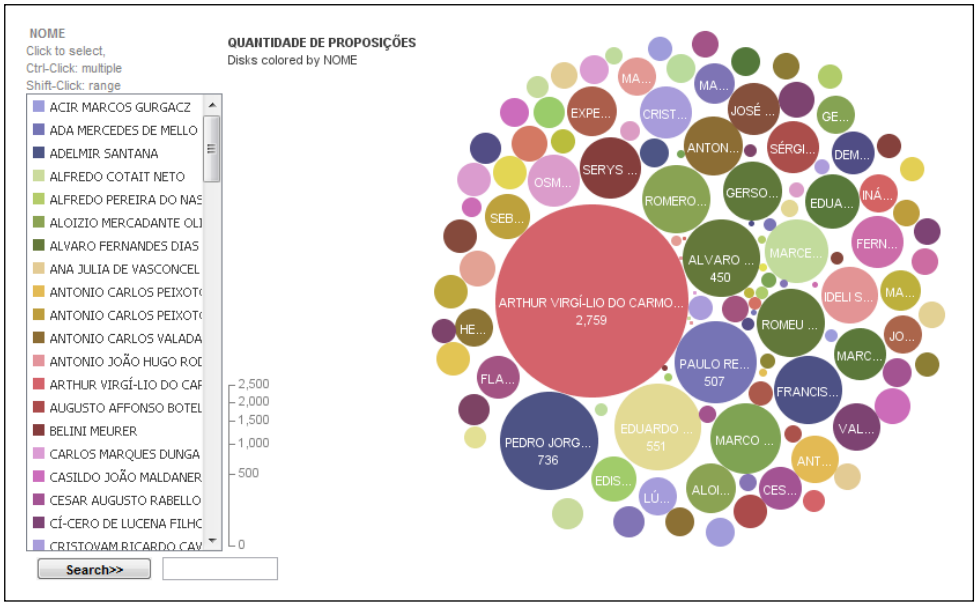


Figura 4.12: Gráfico da quantidade de proposições de políticos cadastrados

Outro exemplo é mostrado na figura 4.13, que apresenta o gráfico da quantidade de políticos cadastrados por grau de instrução.

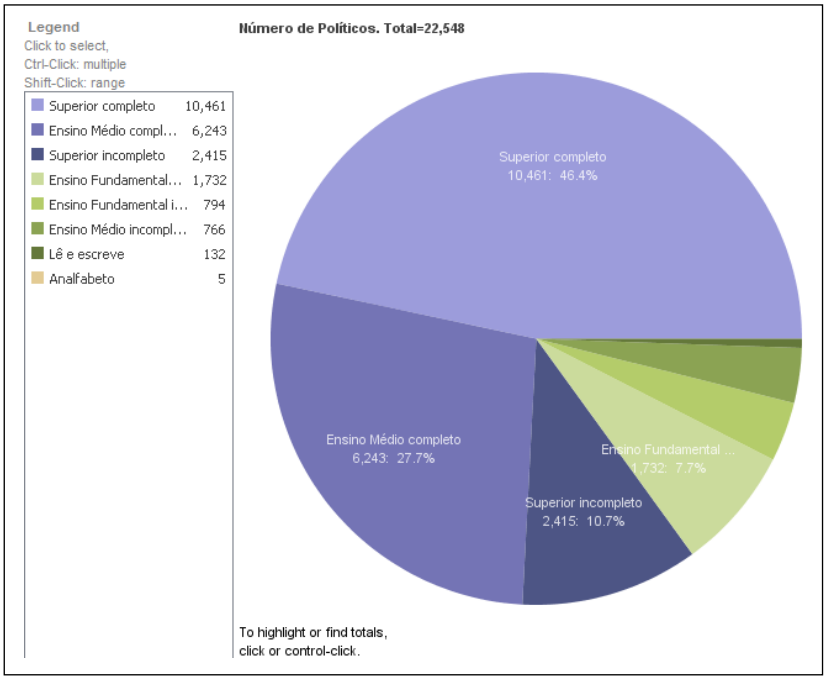


Figura 4.13: Gráfico da quantidade de políticos cadastrados por grau de instrução

Foram geradas também saídas com os dados que, juntamente com a utilização da ferramenta, podem gerar gráficos como as principais palavras utilizadas nos resumos de pronunciamento de políticos cadastrados, conforme é apresentado na figura 4.14.

Figura 4.14: Nuvem de palavras dos resumos dos pronunciamentos dos políticos

4.3. Análise

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, mas esse princípio deve 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 detêm o controle exclusivo dos dados; e possuem licenças livres.

No que diz respeito aos Dados Ligados, o projeto também atende os seus 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.

Além dos princípios básicos que definem Dados Governamentais Abertos e Dados Ligados, buscou-se também seguir as práticas para a publicação dos dados expostas nos capítulos anteriores. Trabalhos futuros para adequar o projeto a todas as práticas definidas são apresentados no capítulo final.

4.4. Considerações Finais

Neste capítulo foi apresentado o projeto de publicação de dados de políticos brasileiros na *Web*, mostrando como ele foi construído e como ele pode ser utilizado.

O capítulo seguinte apresenta as considerações finais do trabalho como um todo, tomando como base os resultados alcançados com o projeto através da utilização dos conceitos apresentados e expondo também propostas de trabalhos futuros.

5. Considerações Finais

Através da realização deste trabalho foi possível perceber a importância da publicação de Dados Governamentais Abertos e das práticas de Dados Ligados na *Web* atual.

Dados governamentais publicados na *Web*, por si só, já possuem um grande valor, pois contribuem para uma maior transparência de informações. A disponibilização dessas informações em formatos abertos e acessíveis permite que elas sejam reutilizadas e misturadas 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 garante que as informações sejam representadas com significado, permite que os dados sejam legíveis por máquinas, proporciona um mecanismo de acesso único e padronizado, facilita a descoberta e o consumo dos dados e permite 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 muito 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.

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 pelo governo. Ao mesmo tempo, mais pessoas e organizações devem publicar dados governamentais por conta própria.

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 em cima desses dados.

Ao disponibilizar dados de políticos brasileiros, poderia ser divulgada juntamente com os dados uma chave primária, como CPF ou título de eleitor, para garantir a identificação e integração das informações.

No que diz respeito aos Dados Ligados, muito se tem sido feito atualmente, mas é preciso difundir ainda mais os conceitos e as práticas de Dados Ligados, bem como criar e melhorar os serviços para facilitar a publicação desses dados, construir mais aplicações e aprimorar as já existentes.

Existe um limitado número de *data sets* publicados se comparados a 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 numero de *data sets* interligados e a quantidade de dados úteis na LOD *Cloud*.

É preciso também melhorar o apoio a infra-estrutura técnica para a publicação de Dados Ligados. Trabalhos e exemplos práticos que tratem do tema de forma mais específica devem ser elaborados de forma a auxiliar neste processo.

Atualmente, existem serviços que auxiliam na descoberta de termos de vocabulários existentes para apoiar a reutilização, mas estes serviços são limitados no sentido de não proverem uma indicação dos melhores vocabulários que podem ser utilizados. Podem ser criados mecanismos que apresentam de forma clara os termos mais utilizados e bem documentados.

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*. Seria interessante também um serviço que fornecesse as principais ligações que podem ser realizadas de acordo com o domínio do *data set*.

Sobre o projeto de Publicação de Dados Ligados na Web, que deu origem ao *data set* “Ligado nos Políticos”, podemos dizer que os dados publicados atendem de maneira geral os princípios básicos dos Dados Governamentais Abertos e dos Dados Ligados. Buscou-se também seguir as práticas de publicação aconselhadas, porém trabalhos futuros devem ser realizados para adequar o projeto a todas as práticas apresentadas.

Em primeiro lugar, é preciso garantir que os dados publicados sejam úteis e atuais. Para isso, é preciso atualizar os dados constantemente através dos *scripts* gerados

para a raspagem de dados. Além disso, mais dados podem ser coletados, como votos, estatísticas, notícias, opiniões sobre o trabalho dos políticos, entre outros. Pode-se também aumentar a quantidade de políticos cadastrados, utilizando dados de candidatos, senadores e deputados de outros anos.

É preciso também definir mais *links* RDF para relacionar os recursos do projeto com outras fontes de dados, através de pesquisas mais detalhadas em outros *data sets*. Além disso, deve-se entrar em contato com proprietários de outros *data sets* relacionados para que eles também definam *links* RDF para o *data set*.

Seria interessante também publicar os dados em outros formatos, além de fornecer um *endpoint* SPARQL para que consultas possam ser realizadas diretamente sobre os dados.

Mais pesquisas em vocabulários existentes podem ser realizadas para garantir a reutilização dos termos. Para os novos termos criados podem ser fornecidas mais informações adicionais e mapeamentos para outros termos.

Pode-se também estudar a possibilidade de usar o nome e o sobrenome do político no URI para identificar os recursos ao invés da chave do político, abrindo uma pagina de desambiguação para registros iguais.

Para melhorar a experiência do usuário, é preciso aprimorar a representação HTML. Além disso, novos conhecimentos sobre os dados coletados podem ser gerados através de novos gráficos e gráficos individuais para cada político cadastrado.

Como outras propostas de trabalhos futuros, pode-se também melhorar a documentação para que as pessoas possam descobrir mais facilmente o que foi publicado, organizar os dados do catálogo usando formatos como o RSS para facilitar e agilizar a divulgação e documentar mecanismos de busca e métodos *RESTful* de obtenção dos dados. Segundo Bizer *et al.* (2009), uma discussão maior em torno de interfaces *RESTful* para sistemas baseados em RDF ainda deve ser realizada.

Assim como a *Web* provocou uma revolução na publicação e no consumo de documentos, Dados Ligados tem o potencial para revolucionar a forma como os dados são acessados e utilizados. E da mesma forma como o Governo Eletrônico revolucionou a relação entre os cidadãos e o governo, Dados Governamentais Abertos tem o potencial de aprimorar e estreitar ainda mais essa relação. Se os desafios ainda existentes forem adequadamente tratados, essas práticas permitirão uma evolução ainda maior na forma como a *Web* é utilizada atualmente.

Referências Bibliográficas

ADIDA, B. et al.. **RDFa in XHTML: Syntax and Processing. A collection of attributes and processing rules for extending XHTML to support RDF.** W3C Recommendation. Outubro, 2008. Disponível em: <http://www.w3.org/TR/rdfa-syntax/>. Último acesso em: Dezembro de 2010.

AGUNE, R. M. FILHO, A. S. G. BOLLIGER, S. P. **Governo Aberto SP: Disponibilização de Bases de Dados e Informações em Formato Aberto**. III Congresso Consad de Gestão Pública. Painel 13/050. 2009. Disponível em: <http://www.repositorio.seap.pr.gov.br/.../governo_aberto_sp_disponibilizacao_de_bases_de_dados_e_informacoes_em_formato_aberto.pdf>. Último acesso em: Dezembro de 2010.

AIRES, J. R. **A democracia digital possível.** Revista Seqüência, no 52, p. 85-104. Julho, 2006. Disponível em: <http://www.buscalegis.ufsc.br/revistas/files/journals/2/articles/29557/public/29557-29573-1-PB.pdf>. Último acesso em: Dezembro de 2010.

ALEXANDER, K. et al.. **Describing Linked Datasets. On the Design and Usage of voiD, the "Vocabulary Of Interlinked Datasets".** LDOW 2009. Madri, Espanha. Abril, 2009. Disponível em: <http://sw-app.org/pub/ldow09-void.pdf>. Último 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/>. Último acesso em: Dezembro de 2010.

BERNERS-LEE, T. **Linked Data**. Design Issues. Julho, 2006. Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html>. Último 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>. Último acesso em: Dezembro de 2010.

Berners-Lee, T. et al. **Uniform Resource Identifiers (URI): Generic Syntax**. Network Working Group. Agosto, 1999. Disponível em: <http://www.ietf.org/rfc/rfc2396.txt>. Último 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/>. Último acesso em: Dezembro de 2010.

BIZER, C. et al.. **Linked Data on the Web**. LDOW 2008. Workshop at the 17th International World Wide Web Conference. Beijing, China. Abril, 2008. Disponível em: <http://events.linkeddata.org/ldow2008/papers/00-bizer-heath-ldow2008-intro.pdf>. Último 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>. Último acesso em: Dezembro de 2010.

BREITMAN, K. **Web Semântica: a Internet do futuro.** Rio de Janeiro: LTC Editora, 2005. 190p.

BRICKLEY, D. GUHA, R.V. **RDF Vocabulary Description Language 1.0: RDF Schema**. W3C Recommendation. Fevereiro, 2004. Disponível em: < http://www.w3.org/TR/rdf-schema/>. Último acesso em: Dezembro de 2010.

COMITÊ GESTOR DA INTERNET NO BRASIL (CGI). **Pesquisa sobre o uso das Tecnologias da Informação e da Comunicação no Brasil 2009**. Núcleo de Informação e Coordenação do Ponto BR. São Paulo, 2010. Disponível em: < http://www.cetic.br/tic/2009/index.htm>. Último 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/>. Último acesso em: Dezembro de 2010.

DINIZ, V. **Como conseguir Dados Governamentais Abertos**. III Congresso Consad de Gestão Pública. Painel 13/049. 2009. Disponível em: <http://www.consad.org.br/sites/1500/1504/00001870.pdf>. Último acesso em: Dezembro de 2010.

EAVES, D. **The Three Laws of Open Government Data**. Conference for Parliamentarians: Transparency in the Digital Era. Setembro, 2009. Disponível em: <http://eaves.ca/2009/09/30/three-law-of-open-government-data/>. Último acesso em: Dezembro de 2010.

FIELDING, R. T. **Architectural Styles and the Design of Network-based Software Architectures**. University of California, Irvine. 2000. Disponível em: < http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>. Último acesso em: Dezembro de 2010.

FIELDING, R. T. **Hypertext Transfer Protocol -- HTTP/1.1**. Network Working Group. The Internet Society. 1999. Disponível em: < http://www.w3.org/Protocols/rfc2616/rfc2616.html>. Último acesso em: Dezembro de 2010.

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

HAUSENBLAS, M. **Exploiting Linked Data For Building Web Applications**. IEEE Internet Computing. 2009a. Disponível em: <http://sw-app.org/pub/exploit-lod-webapps-IEEEIC-preprint.pdf>. Último acesso em: Dezembro de 2010.

HAUSENBLAS, M. **Linked Data Applications.** DERI Technical Report. DIGITAL ENTERPRISE RESEARCH INSTITUTE. Galway, Irlanda. Julho, 2009b. Disponível em: <http://linkeddata.deri.ie/sites/linkeddata.deri.ie/files/lod-app-tr-2009-07-26_0.pdf> Último acesso em: Dezembro de 2010.

HEATH, T. **An Introduction to Linked Data.** Platform Division. Talis Information Ltd. Austin, Texas. Fevereiro, 2009. Disponível em: <http://tomheath.com/slides/2009-02-austin-linkeddata-tutorial.pdf>. Último acesso em: Dezembro de 2010.

HEATH, T. et al.. **How to Publish Linked Data on the Web**. Half-day Tutorial at ISWC2008. Karlsruhe, Alemanha. Outubro, 2008. Disponível em: <http://events.linkeddata.org/iswc2008tutorial/how-to-publish-linked-data-iswc2008-slides.pdf>. Último acesso em: Dezembro de 2010.

MACMANUS, R. **The State of Linked Data in 2010**. Publicado em: <http://www.readwriteweb.com/>, 2010. Disponível em: <http://www.readwriteweb.com/archives/the_state_of_linked_data_in_2010.php>. Último acesso em: Dezembro de 2010.

OPENGOVDATA.ORG. **Open Government Data Principles.** California. Dezembro, 2007. Disponível em: <http://resource.org/8_principles.html>. Último acesso em: Dezembro de 2010.

PRUD'HOMMEAUX, E. SEABORNE, A. **SPARQL Query Language for RDF**. W3C Recommendation. Janeiro, 2008. Disponível em: <http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/>. Último acesso em: Dezembro de 2010.
MCGUINNESS, D. L. VAN HARMELEN, F. **OWL Web Ontology Language Overview**. W3C Recommendation. Fevereiro, 2004. Disponível em: <http://www.w3.org/TR/owl-features/>. Último acesso em: Dezembro de 2010.

NOTTINGHAM, M. SAYRE, R. **The Atom Syndication Format**. Network Working Group. Atom Enabled. Dezembro, 2005. Disponível em: < http://xml.resource.org/public/rfc/html/rfc4287.html>. Último acesso em: Dezembro de 2010.

RDF WORKING GROUP. **Resource Description Framework (RDF)**. Fevereiro, 2004. Disponível em: <http://www.w3.org/RDF/>. Último acesso em: Dezembro de 2010.

RSS 2.0 SPECIFICATION. **RSS 2.0 at Harvard Law. Internet technology hosted by Berkman Center**. Julho, 2003. Disponível em: <http://cyber.law.harvard.edu/rss/rss.html>. Último 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>. Último acesso em: Dezembro de 2010.

STANKOVIC, M. et al.. **Looking for Experts? What can Linked Data do for You?** LDOW 2010. Raleigh, Estados Unidos. Abril, 2010. Disponível em: <events.linkeddata.org/ldow2010/papers/ldow2010_paper19.pdf>. Último acesso em: Dezembro de 2010.

TENNISON, J. SHERIDAN, J. **SemWebbing the London Gazette**. 2008. Disponível em: <http://assets.expectnation.com/15/event/3/SemWebbing%20the%20London%20Gazette%20Paper%201.pdf>. Último acesso em: Dezembro de 2010.

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>. Último acesso em: Dezembro de 2010.

W3C ESCRITÓRIO BRASIL. **O Governo de inovação na Copa 2014: uso de redes sociais e dados abertos.** II Seminário de Inovação em Governo Eletrônico. Porto Alegre, Rio Grande do Sul. Setembro, 2010. Disponível em: <http://www.procergs.rs.gov.br/uploads/1285856001W3C_Seminario_Inovacao_eGov_POA_17092010.pdf>. Último acesso em: Dezembro de 2010.