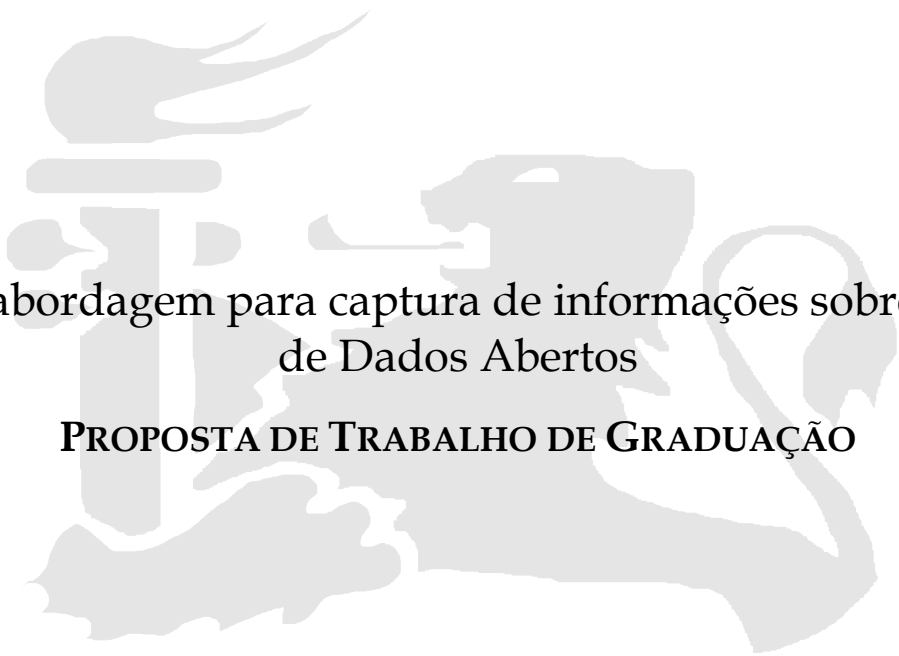




UNIVERSIDADE FEDERAL DE PERNAMBUCO
BACHARELADO EM SISTEMAS DE INFORMAÇÃO
CENTRO DE INFORMÁTICA
2014.1



Uma abordagem para captura de informações sobre o uso
de Dados Abertos

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Lairson Emanuel R. de Alencar Oliveira (lerao@cin.ufpe.br)

Orientadora: Bernadette Farias Lóscio (bfl@cin.ufpe.br)

Recife, 05 de maio de 2014

1. Contextualização

Um dos fatores que mais contribuem para o avanço da interação entre governo e sociedade é a Internet. Somada à Internet, temos as tecnologias da informação e comunicação (TICs), as quais também contribuem significativamente para a formação de um cenário que facilita tal interação, ampliando o acesso e a utilização dos dados públicos pelos cidadãos [1]. Com a aplicação de técnicas e ferramentas adequadas, torna-se possível gerar novos conhecimentos, informações úteis e melhores visualização dos dados públicos governamentais [2].

Segundo Gusmão [1], além de aproximar o cidadão do Estado, a disponibilização de dados abertos fortalece a democracia e o acesso à informação pública, que é uma prerrogativa de todo cidadão. Sendo, portanto, a possibilidade de reusar, recombina e dar a tais dados novos significados, uma ampliação necessária desse direito.

A *Open Knowledge Foundation (OKF)*, em síntese, define que dados são abertos quando qualquer pessoa pode livremente usá-los, reutilizá-los e redistribuí-los, estando sujeito a, no máximo, a exigência de creditar a sua autoria e compartilhar pela mesma licença [3]. A Cartilha Técnica para Publicação de Dados Abertos no Brasil [4] acrescenta que Dados Abertos constituem a publicação e disseminação de dados e informações públicas na Web, seguindo alguns critérios que possibilitam sua reutilização e o desenvolvimento de aplicativos por toda a sociedade. Além disso, afirma que o paradigma de dados abertos está fundamentado na constatação de que o dado, quando compartilhado abertamente, tem seu valor e seu uso potencializados.

Sancionada em 18 de novembro de 2011, a Lei de Acesso à Informação Pública, começou a vigorar em 16 de maio de 2012 no Brasil, onde foi permitido a qualquer pessoa ter acesso a documentos e informações que estejam sob a guarda de órgãos públicos, em todos os poderes e níveis de governo [5]. Ainda em 2012, foi instituído a Infraestrutura Nacional de Dados Abertos (INDA), que é um conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas no modelo de Dados Abertos [6].

Em dezembro de 2011, em paralelo aos esforços para constituir a INDA, foi lançado o Portal Brasileiro de Dados Abertos, o dados.gov.br (que também é acessível por data.gov.br). O portal é a ferramenta disponibilizada pelo governo que agrega e permite o uso dos dados e informações públicas. Prezando pela simplicidade, o portal tem o

objetivo de promover a interlocução entre atores da sociedade e o governo, a fim de pensar a melhor utilização dos dados em prol de uma sociedade melhor [7]. Além disso, a exemplo de outros portais similares, visa facilitar o acesso às bases de dados públicas, permitindo a localização dos dados em formato aberto a partir de um único lugar e por meio de uma ferramenta de busca. Hoje, o portal brasileiro reúne 112 conjuntos de dados [7].

O portal de dados do governo brasileiro utiliza o software CKAN¹, um software *open source* que também é usado por vários outros países, como Uruguai, Canadá, Reino Unido, Estados Unidos, Irlanda, Romênia, Austrália, Holanda, Itália, Argentina e Alemanha. O CKAN é uma ferramenta para catalogação de dados, desenvolvido pela OKF com o objetivo de permitir uma rápida publicação, uso e compartilhamento de dados, além de ser extensível e personalizável [9].

Para descrever catálogos e conjuntos de dados, o W3C recomenda o *Data Catalog Vocabulary* (DCAT), que é um vocabulário projetado para facilitar a interoperabilidade entre os catálogos de dados publicados na web. Assim, além de facilitar a descrição, também permite que aplicativos consumam os metadados de vários catálogos [8].

Contudo, ainda é um grande desafio rastrear o uso dos conjuntos de dados e as aplicações que fazem uso desses dados. De uma maneira geral, busca-se capturar os dados que são usados em um aplicativo, uma descrição do que o aplicativo faz e qual o problema que ajuda a resolver. Da mesma forma, espera-se obter um retorno sobre possíveis melhorias nos conjuntos de dados, quais dados do conjunto são mais utilizados, se os dados originais sofreram transformações ou se podem ser melhorados. Assim, um vocabulário voltado para a descrição do uso dos dados e a implementação de uma ferramenta que capture essas informações tornam-se necessários. Vale salientar que tais informações sobre o uso dos dados, bem como feedback sobre os conjuntos de dados, podem ser de grande utilidade na seleção dos dados a serem utilizados, além de contribuir para a melhoria da qualidade dos dados [10].

¹ ckan.org

2. Objetivos

O objetivo deste trabalho é desenvolver uma abordagem para captura de informações sobre o uso de dados abertos. Como parte da abordagem serão desenvolvidos um vocabulário para descrição de informações relevantes sobre o uso de dados e uma extensão para a ferramenta CKAN, que permita a captura de tais informações. Além disso, por meio da extensão proposta será possível fazer a ligação entre a aplicação e os conjuntos de dados utilizados para seu desenvolvimento. O intuito é a partir do cadastro das aplicações, criar um repositório central e rastrear o uso dos conjuntos de dados. Assim, será possível obter um feedback sobre possíveis melhorias nos dados, quais dados do conjunto estão sendo mais utilizados e se houve transformação nos dados originais.

3. Cronograma

O cronograma de atividades previsto para o desenvolvimento deste Trabalho de Graduação é apresentado na tabela a seguir (Tabela 1), onde o período de cada mês está dividido em colunas que representam, aproximadamente, o período de uma semana.

Atividade	Maio				Junho				Julho				Agosto			
Levantamento do estado da arte e definição do escopo	X	X	X	X												
Estudo dos trabalhos relacionados e ferramentas associadas		X	X	X												
Proposta de vocabulário para descrição de informações relevantes sobre o uso de dados					X	X										
Implementação de extensão para o CKAN						X	X	X	X	X						
Validação e testes									X	X	X					
Elaboração do relatório final									X	X	X	X	X			
Preparação e defesa														X	X	

Tabela 1: Cronograma de atividades

4. Possíveis Avaliadores

Os possíveis avaliadores desse trabalho de graduação serão:

- Kiev Santos da Gama
- Bernadette Farias Lóscio

5. Referências

[1] **Dados abertos para a Democracia na Era Digital**. – Brasília: Fundação Alexandre de Gusmão, 2011. 84 p.

[2] Moreira, G. A; Lóscio, B. F. **GovPEX3: uma ferramenta para publicação das despesas do Governo de Pernambuco em RDF**. Recife, 2013. Disponível em: <<http://www.cin.ufpe.br/~tg/2012-2/gam.pdf>>. Acesso em: 28 de abril de 2014.

[3] Open Knowledge Foundation. **What is Open?** Disponível em: <<https://okfn.org/opendata>>. Acesso em: 28 de abril de 2014.

[4] Secretaria de Logística e Tecnologia da Informação – SLTI; Ministério do Planejamento Orçamento e Gestão – MP. **Cartilha Técnica para Publicação de Dados Abertos no Brasil**. Disponível em: <<http://dados.gov.br/cartilha-publicacao-dados-abertos>>. Acesso em: 28 de abril 2014.

[5] CGU. **Lei de Acesso à Informação entra em vigor hoje**. Disponível em: <<http://www.acessoainformacao.gov.br/acessoainformacaogov/noticias/0092012.asp>>. Acesso em: 28 de abril de 2014.

[6] Infraestrutura Nacional de Dados Abertos - INDA. Disponível em: <<http://wiki.gtinda.ibge.gov.br/MainPage.ashx>>. Acesso em: 28 de abril de 2014.

[7] Portal Brasileiro de Dados Abertos. **Sobre o dados.gov.br**. Disponível em: <<http://dados.gov.br/sobre>>. Acesso em: 28 de abril de 2014.

[8] W3C Recommendation. **Data Catalog Vocabulary (DCAT)**. Disponível em: <<http://www.w3.org/TR/vocab-dcat>>. Acesso em: 28 de abril de 2014.

[9] CKAN. **About**. Disponível em: <<http://ckan.org/about/>>. Acesso em: 28 de abril de 2014.

[10] W3C. **Data usage notes**. Disponível em: <https://www.w3.org/2013/dwbp/wiki/Data_usage_notes>. Acesso em: 28 de abril de 2014.

6. Assinaturas

Lairson Emanuel R. de Alencar Oliveira
Orientando

Bernadette Farias Lóscio
Orientadora

Recife, 05 de maio de 2014