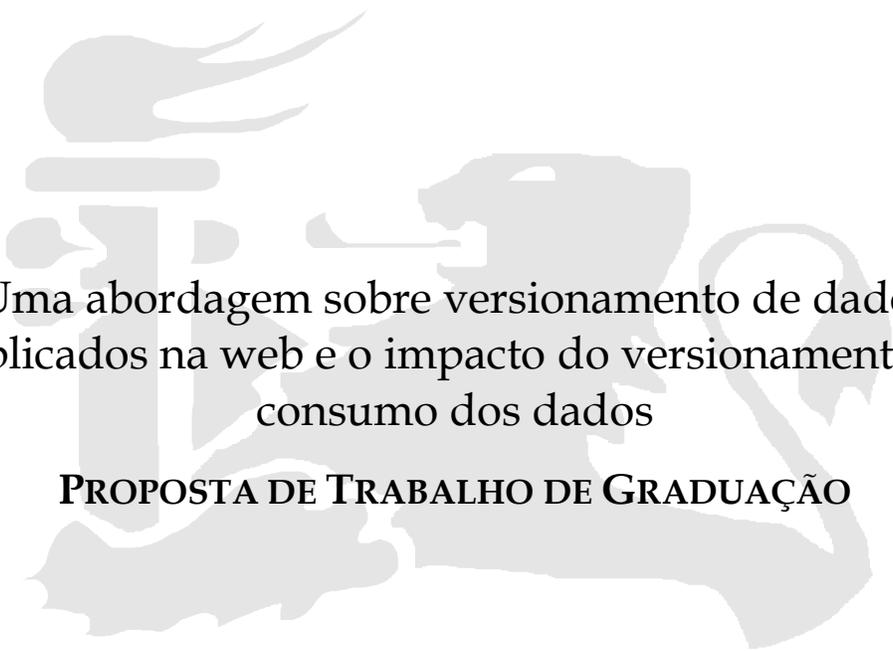




UNIVERSIDADE FEDERAL DE PERNAMBUCO
BACHARELADO EM SISTEMAS DE INFORMAÇÃO
CENTRO DE INFORMÁTICA
2016.2



Uma abordagem sobre versionamento de dados
publicados na web e o impacto do versionamento no
consumo dos dados

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Maurício Taumaturgo (mto@cin.ufpe.br)

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

Recife, 19 de setembro de 2016

1. Contextualização

Por mais de 50 anos, o acesso à água potável da cidade de Zanesville, Estados Unidos, foi disponibilizado para diversos bairros do condado de Muskingum em Ohio, Estados Unidos. Porém, moradores de uma região predominantemente negra da cidade de Zanesville tinha acesso apenas à água contaminada da chuva ou teria que buscar na torre de água mais próxima. Esse caso foi parar na justiça e, durante a sua defesa, uma das principais evidências utilizadas foi um mapa construído, a partir de dados publicados na Web, mostrando a relação entre as casas ocupadas por moradores brancos e as casas ligadas a tubulação de água da cidade. A justiça decidiu em favor da população da área predominante negra chegando a um acordo de \$10.9 milhões de dólares [1]. Situações como essa, demonstram que a publicação de dados na Web tem, cada vez mais, elevado impacto na forma como as cidades são geridas e como as políticas públicas são criadas e implementadas. Segundo Simon Chignard [2], o movimento de dados abertos tem alcançado um número significativo de governos ao redor do mundo. Este movimento tem impactado a sociedade de diversas maneiras como, por exemplo, na redefinição da política de distribuição da água no condado de Muskingum, citado anteriormente.

Segundo o Open Definition [3], *“dados e conteúdos abertos podem ser usados, modificados e compartilhados livremente por qualquer um para qualquer propósito”*. Neste contexto, governos ao redor do mundo estão cada vez mais aderindo à publicação, na Web, de dados da população. Dados estes que, quando publicados na internet, se tornam acessíveis por qualquer um ao redor do mundo, podendo, quando publicados em formato aberto, inclusive ser modificado por todos. Segundo dados do Global Open Data Index [4], mais de 120 países já publicaram algum conjunto de dados (dataset) na Web. De acordo com o relatório Open Data Impact [1], a publicação desses dados impacta em diversas dimensões, tais como crescimento econômico, inovação, tomada de decisões, mobilização social, entre outras. Portanto, há de se concluir que a publicação de conjuntos de dados na Web, principalmente governamentais, é uma tendência mundial com grandes impactos para a sociedade. Diante desta tendência, é de extrema importância que os conjuntos de dados publicados na Web estejam sempre atualizados, a fim de garantir a sua utilidade, bem como a confiança daqueles interessados em consumir o dataset. De acordo com o documento Data on the Web Best Practices (DWBP) [6], da

W3C, os conjuntos de dados podem ser atualizados de acordo com um tempo específico, outros são atualizados à medida que algumas melhorias e/ou correções são realizadas.

Diante do crescente número de datasets publicados e da facilidade com que estes dados são modificados, faz-se necessário utilizar meios capazes de controlar as diferentes versões e atualizações nos conjuntos de dados. De acordo com Chacon e Straub (2015, p. 27), controle de versão é *“um sistema que registra as mudanças feitas em um arquivo ou um conjunto de arquivos ao longo do tempo de forma que você possa recuperar versões específicas”* [5].

Conforme o documento DWBP, o versionamento de conjuntos de dados possibilita diversos entendimentos, entre eles determinar a versão do dataset que está sendo utilizada, determinar também se há uma nova versão disponível e entender como uma versão do dataset difere da outra. Por outro lado, ainda conforme o DWBP, não há um consenso sobre quando um conjunto de dados é considerado uma versão atualizada ou se é um novo conjuntos de dados. Assim como, também não há um consenso sobre como versionar estes conjuntos de dados de forma que qualquer indivíduo que consumir estes dados entenda quais mudanças aconteceram ao comparar diversas versões.

Neste contexto, é de extrema importância o desenvolvimento de soluções que permitam o versionamento, de forma adequada, de conjuntos de dados publicados na Web. Tal solução deve ser de âmbito global e deverá ser possível aplicá-la para qualquer conjunto de dados. Este versionamento, impactará os consumidores dos dados de diversas maneiras, possibilitando assim o entendimento a respeito do conjunto de dados que está sendo utilizado e suas diversas versões. Desta maneira, faz-se necessário entender o impacto que o versionamento de dados publicados na Web tem para os consumidores dos dados, bem como desenvolver uma solução capaz de padronizar este versionamento.

2. Objetivos

Este trabalho tem como objetivo geral propor uma solução para o problema de versionamento de conjuntos de dados publicados na Web, bem como a análise do impacto que este versionamento tem para os consumidores de dados. Para tal, os diferentes padrões de versionamento de software serão utilizados como referência para o desenvolvimento do novo modelo aplicado ao contexto de dados na Web. Além disso, serão analisados, junto aos consumidores dos dados, quais os impactos que o versionamento dos dados tem na construção de aplicações que fazem uso dos dados publicados na Web.

3. Cronograma

A tabela abaixo apresenta o cronograma das atividades que serão desenvolvidas na realização deste Trabalho de Graduação.

Cronograma													
Atividade	Setembro			Outubro			Novembro			Dezembro			
Análise do contexto	x												
Identificação do problema	x	x											
Definição do Objetivo		x											
Estado da Arte e Referências			x	x	x								
Definição do Escopo				x									
Pesquisa sobre Trabalhos Relacionados					x								
Desenvolvimento do Padrão para Versionamento de Dados Publicados na Web						x	x	x					
Validação do Padrão Desenvolvido							x	x					
Análise do impacto do Versionamento de Dados Publicados na Web									x	x			
Finalização do Documento de TG										x	x		
Preparação da Defesa												x	x

Tabela 1: Cronograma de atividades

4. Possíveis Avaliadores

Os nomes abaixo referem-se aos possíveis avaliadores deste trabalho.

- Kiev Santos da Gama – Cin UFPE
- Bernadette Farias Lóscio – Cin UFPE

5. Referências

- [1] Verhulst S, Young A. Open data impact. Disponível em <http://odimpact.org/static/files/open-data-impact-key-findings.pdf>. Acesso em 17 de setembro de 2016.
- [2] Chignard S. A brief history of Open Data. Disponível em <http://www.paristechreview.com/2013/03/29/brief-history-open-data/>. Acesso em 17 de setembro de 2016.
- [3] The Open Definition. Disponível em <http://opendefinition.org>. Acesso em 17 de setembro de 2016.
- [4] Global Open Data Index (2014). Disponível em <http://index.okfn.org/place/>. Acesso em 17 de setembro de 2016
- [5] Chacon S, Straub B. Pro Git. 2. Ed. 2016. Disponível em <https://progit2.s3.amazonaws.com/en/2016-03-22-f3531/progit-en.1084.pdf>. Acesso em 17 de setembro de 2016.
- [6] W3C Candidate Recommendation. Data on the Web Best Practices. Disponível em <http://www.w3.org/TR/dwbp/>. Acesso em 17 de setembro de 2016.