



Estudo comparativo de datasets governamentais do Brasil e da Colômbia, com dados de agricultura e desenvolvimento rural

Jacquelin Camperos Reyes^a

Transcrição da apresentação

Olá pessoal bom dia. Grata com os organizadores do oitavo CEDIT, Ciclo de Estudos de Dados, Informação e Tecnologia, evento organizado e que acontece na cidade de Tupã na Universidade Estadual Paulista UNESP. Nessa ocasião vou apresentar o resultado do estudo que foi realizado em datasets publicados pelos governos do Brasil e da Colômbia especificamente nos assuntos de Agricultura e Desenvolvimento Rural. O foco da pesquisa está na aderência desses datasets a princípios de dados abertos publicados por governos.

Estudo Comparativo de datasets governamentais do Brasil e da Colômbia com dados de Agricultura e Desenvolvimento Rural.

Os autores são, quem fala, a Jacquelin Teresa Camperos Reyes, doutoranda do Programa de Pós-Graduação em Ciência da Informação da Unesp, campus de Marília, o professor Ricardo

^a Mestre em Ciência da Informação. Estudante de Doutorado em Ciência da Informação na UNESP - Universidade Estadual Paulista. E-mail: jtcamperos@hotmail.com. ORCID: <https://orcid.org/0000-0002-0078-5376>. Currículo: <http://lattes.cnpq.br/5415219564237576>.

Sant'Ana e o professor José Eduardo Santarém Segundo.

A observância, como disse, dos datasets, é focando princípios que foram criados na Open Government Data no ano 2007, esses oito princípios que vão ser observados nesse estudo são: 1º completo, se os dados são todos os que os governos produzem nos seus ambientes informacionais; 2º, primário, quer dizer acerca da granulariedade desses dados: conforme eles são produzidos não deveriam ter nenhum tipo de agregação nem transformações; 3º oportuno, esses dados são oportunos na medida de que tão rápido eles estão sendo publicados para preservar o seu valor temporal; 4º acessível: qual a gama de usuários e propósitos que podem utilizar esses dados que os governos estão publicando; 5º processado por máquina: se esses dados estão sendo publicados de forma estruturada para permitir o processamento automatizado; 6º não discriminatório, disse acerca de não requerer registro por parte dos usuários de tal forma que usuários anônimos deveriam ser permitidos; 7º não proprietário, tem a ver com que não devem ter controle exclusivo por alguma entidade ou seja quem não tem um aplicativo específico que seja o único no qual possa ser usados se esses dados; finalmente, 8º Livre-licença, que não estão estejam sujeitos a direito autoral nem marca registrada.

Resultados da pesquisa: no contexto do Brasil foram recuperados 32 datasets e na Colômbia 70, lembrando que a pesquisa foi desenvolvida no mês de agosto do ano de 2016.

Com relação aos princípios 1, 2 e 3, completo, primário é oportuno, o resultado da pesquisa disse que há ausência de critérios específicos que permitam mensurar esses princípios. Porque? Porque teriam que estar disponibilizados no site de dados alguns inventários para que os usuários possam ver se tudo o que produz esse órgão de governo está sendo publicado, não há um ponto de comparação. Para os primários, o usuário não sabe o estado original desses dados na fonte para inferir sobre a

granulariedade dos dados; se são oportunos ou não, também não é possível observar, porque essas datas nas que são produzidos os dados dependem de contextos internos de cada órgão, de cada instituição, em cada país. Já no princípio 4, se eles são acessíveis, foi constatado que 100% dos datasets não precisam nenhum tipo de identificação em nenhum dos dois países; também não perguntam acerca do uso futuro desses dados, então não há restrição em quanto aos sistemas operativos nem outros aspectos técnicos para esses usuários que estão acessando os dados. No Brasil todos os datasets estão publicados em um formato estruturado como o XML o que permite um tipo de manuseio desses dados, tem outros formatos além do XML, mas destaca-se esse pelas possibilidades de processamento. Já no caso da Colômbia XML, RDF, JSON e CSV são formatos que estão no 80% dos datasets, porém, existem alguns formatos que não são processáveis por máquina e que estão sendo utilizados em esse site de governo, como são os arquivos com compressão e arquivos que são somente para visualização gráfica.

Comparando os dois países, se observa então que na maioria dos casos, os datasets, publicam-se em formatos que permitem um tipo de processamento automatizado.

No caso do Princípio 6 observou-se que eles não têm restrições discriminatórias no datasets analisados, não requerem registro dos usuários, então permitem buscas anônimas.

No princípio 7, no Brasil todos estão publicados, os datasets que foram recuperados, em um formato não proprietário, não há um requisito de algum software específico que requeira custo ou pagamento. Já no caso da Colômbia 10% desses datasets se apresentam no formato Excel que é desenvolvido pela multinacional Microsoft. Esse formato é um formato que não cumpre com esse princípio 7 de não ser um formato proprietário e esse formato está sendo utilizado nesse país não como uma alternativa, então o usuário ou utiliza esse arquivo como está

sendo publicado e ele poderá transformá-lo em outro não proprietário, mas, ele não está sendo publicado como uma opção.

Finalmente no princípio 8 encontram-se vários tipos de características relacionadas com que eles sejam livre e dê licença, porém, não existe uma documentação adequada para que o usuário desses datasets saiba especificamente quais são as suas implicações legais, nesses tipos de datasets que foram recuperados.

Finalmente já como considerações finais, esses dois países satisfazem nas recomendações dos princípios 4 e 6 permitindo o uso não discriminado de usuários e também amplo para características tecnológicas.

No caso da Colômbia, ela tem um maior número de datasets, porém, tem que trabalhar no fato de que todos esses dados estejam publicados em formatos estruturados processáveis por máquina e em arquivos não proprietários, assim como no Brasil que o 100% dos datasets têm essas duas características para trabalhar em esses princípios 5 e 7.

É bem importante pesquisar e continuar pesquisando em torno da publicação dos dados abertos dos países, especificamente nos setores acadêmicos, mas, dentro de setores econômicos que são tão representativos para economias que são baseadas numa grande porcentagem pelas atividades especificamente de Agricultura e Desenvolvimento Rural. Porque? Para melhorar a qualidade desses frutos informacional e portanto no fortalecimento da relação do estado-cidadão.

Deixo aqui os contatos dos autores, fico grata pela atenção de vocês e convido a continuar acompanhando as publicações e os eventos organizados pela Unesp de Tupã e pelo Grupo de Pesquisa Novas Tecnologias de Acesso a Dados e no CEDIT. Obrigada.

Vídeo da Apresentação

Título: Estudo comparativo de datasets governamentais do Brasil e da Colômbia, com dados de agricultura e desenvolvimento rural



Disponível em:

http://dadosabertos.info/enhanced_publications/9cedit/video.php?id=4

Slides da Apresentação

Título: Estudo comparativo de datasets governamentais do Brasil e da Colômbia, com dados de agricultura e desenvolvimento rural



Disponível em:

http://dadosabertos.info/enhanced_publications/9cedit/presentation.php?id=4