

INFORMAÇÃO PARA AÇÃO: Desenvolvimento de um Portal de Dados Abertos Sobre Agrotóxicos

Área Temática: Informática Comunitária e Inclusão Digital / Engenharia e Sustentabilidade

Alan Freihof Tygel¹, Leonardo Gonçalves², Mayara Santos³, Gabriel Marques⁴, Maria Luiza Machado Campos⁵

¹Programa de Pós-Graduação em Informática – PPGI/UFRJ – alantygel@ppgi.ufrj.br

² Departamento de Ciência da Computação – DCC/UFRJ – leonardo.goncalves.2711@gmail.com

³ Departamento de Ciência da Computação – DCC/UFRJ – mayara.morais.santos@gmail.com

⁴ Departamento de Ciência da Computação – DCC/UFRJ – gabrielmarques92@hotmail.com

⁵ Programa de Pós-Graduação em Informática – PPGI/UFRJ – mluiza@ppgi.ufrj.br

Resumo

Em 2008, o Brasil se tornou o maior consumidor de agrotóxicos do mundo, posto que mantém ainda em 2014. O uso de mais de 1 bilhão de litros de venenos por ano têm causado grande impacto na saúde e no meio ambiente. Entretanto, a falta de dados confiáveis sobre os vários aspectos do uso de agrotóxicos e seus efeitos dificulta a ação do poder público e dos movimentos sociais na denúncia das intoxicações e anúncio das alternativas. Neste sentido, o Observatório de Atenção Permanente ao Uso de Agrotóxicos foi desenvolvido com objetivo de ser um portal de dados abertos que possa subsidiar ações contra o uso de venenos e em defesa da vida.

Neste projeto, foi implantado o sistema de gestão de conteúdo DKAN, que permite a disponibilização e visualização de conjuntos de dados através de um portal na Internet. Ferramentas que permitem processos interativos de diálogo sobre os dados estão sendo utilizadas, entendendo que a simples disponibilização dos dados não gera mobilização

por si só. Neste sentido, o desafio do Portal é ser um sistema útil para os movimentos sociais que atuam neste campo.

Palavras-chave: Dados abertos, agrotóxicos, informática e sociedade, saúde.

Introdução

O Brasil é o líder do ranking mundial de consumo de agrotóxicos. O uso excessivo de agrotóxicos está diretamente relacionado à atual política agrícola do país, que privilegia o chamado agronegócio em detrimento da agricultura familiar e ecológica. Assim, mais de 823.226 toneladas de venenos foram jogadas nas lavouras somente em 2012, segundo dados do Sindicato Nacional da Indústria de Produtos para Defesa Vegetal⁴⁶, num mercado que movimentou U\$ 9,71 bilhões no mesmo ano. Os impactos na saúde e no meio-ambiente têm sido cada vez mais graves, e estudos científicos têm avançado no sentido de comprovar a ligação entre o uso de agrotóxicos e doenças como câncer, má-formação fetal e depressão, que em muitos casos pode levar ao suicídio (CARNEIRO et al., 2012).

No estado do Rio de Janeiro, a situação não é diferente. Das 105 amostras analisadas pelo Programa de Análise de Resíduo de Agrotóxicos em Alimentos em 2010 – PARA (ANVISA, 2010), 41 (39%) continham mais resíduos de agrotóxicos do que o permitido, e/ou continham agrotóxicos não permitidos. Nacionalmente, apenas em 37% das amostras analisadas não foi detectado resíduo de agrotóxico.

Diante deste cenário, um conjunto de organizações assumiu em 2011 o esforço coletivo de criação da Campanha Permanente Contra os Agrotóxicos e Pela Vida⁴⁷. A campanha visa combater a utilização de agrotóxicos e a ação de empresas associadas a eles (produtoras e comercializadoras), explicitando as contradições geradas pelo modelo de produção agrícola imposto pelo agronegócio. Além disso, propõe um modelo alternativo de produção de alimentos, que não prejudica o meio ambiente, os trabalhadores e nem os consumidores, notadamente através da agroecologia.

No contexto da campanha, tem-se verificado que existe disponível na Internet uma grande quantidade de dados relacionados ao uso de

⁴⁶<http://www.sindiveg.org.br/>

⁴⁷<http://contraosagrototoxicos.org/>

agrotóxicos e aos danos causados por eles. No entanto, esses dados encontram-se dispersos e desconectados. Além disso, os participantes da Campanha sentem a necessidade de um espaço que possa reunir as informações novas geradas por pesquisadores, militantes de movimentos sociais e agricultores, de modo que essa informação seja útil para a ação contra a utilização de agrotóxicos e em defesa da vida.

Assim, este artigo descreve o desenvolvimento de um Portal de Dados Abertos sobre Agrotóxicos⁴⁸, que possa responder a este anseio da comunidade de pesquisadores e movimentos sociais que compõem a Campanha Permanente Contra os Agrotóxicos e Pela Vida.

Dados sobre agrotóxicos

Um portal de dados abertos sobre agrotóxicos no Brasil deve reunir dados que sirvam como subsídio para entender o contexto em que o ocorre a utilização dos venenos no país. Ao refletir sobre os tipos de dados necessários, chega-se rapidamente à conclusão de que áreas de conhecimento tão distintas quanto Saúde Coletiva, Química, Direito, Agricultura, Agronomia, dentre outras, devem se juntar para que se possa ter uma visão que alcance o nível de complexidade do problema. A partir da análise de diversas bases disponíveis na Internet e de entrevistas com especialistas, chegou-se a dez categorias de dados. São elas:

1. **Agravos à Saúde e ao Meio-Ambiente:** Bases de dados contendo ocorrências de saúde e contaminações ambientais em que pelo menos uma das causas pode estar relacionada a agrotóxicos.
2. **Registro de Agrotóxicos:** Bases de dados contendo informações sobre agrotóxicos, relacionadas aos registros nos órgãos competentes.
3. **Utilização:** Bases de dados que registram o consumo de agrotóxicos no país, além dos registros de produção, comercialização, exportação e importação de agrotóxicos no Brasil.
4. **Legislação:** Bases de dados contendo legislações que regulam os agrotóxicos nos diversos âmbitos.
5. **Agroecologia:** Bases de dados que registram experiências de agroecologia, apresentando alternativas de produção de alimentos sem agrotóxicos.

⁴⁸<http://www.contraosagrototoxicos.org/dados-abertos>

6. **Trabalhos Acadêmicos:** Bases de dados que registram trabalhos acadêmicos no campo dos agrotóxicos.
7. **Notícias:** Bases de dados de notícias jornalísticas relacionadas a agrotóxicos e seus efeitos na saúde e no meio-ambiente.
8. **Produção Agrícola:** Bases de dados da produção agrícola brasileira, que possam dar indícios sobre a utilização de determinados venenos em cada região.
9. **Qualidade de vida de uma população:** Índices que procuram medir a qualidade de vida de populações, ponderando diversos aspectos como saúde, educação e renda.

Neste trabalho, foi realizado um levantamento inicial das bases de dados que podem compor o portal em cada uma das categorias descritas acima. Para cada base de dados, foi analisado o seu nível de abertura.

O conceito de dados abertos possui diversas definições. Aqui, usaremos como referência as 5 estrelas propostas por BERNERS-LEE (2014). Esta definição é baseada em critérios legais (dados com licença que permita o uso público), de formato dos dados (devem ser legíveis por máquina) e de metadados. As cinco estrelas são definidas de forma incremental cumulativa, segundo os critérios a seguir:

1. Dados disponíveis na Internet (em qualquer formato) livre de propriedade intelectual;
2. Dados estruturados (não fornecidos como imagens ou tabelas, e sim em uma planilha);
3. Dados representados em formatos não proprietários, como por exemplo CSV ou ODS;
4. Identificação via URIs (Uniform Resource Identification), que permitem identificar unicamente recursos de informação;
5. Dados ligados a outros de bases distintas.

Esta classificação, no entanto, não encontra uma boa aplicação no contexto dos dados analisados. As estrelas 4 e 5 colocam os dados no nível da Web Semântica (ver Seção 3), e ainda há muito poucos exemplos nesse sentido nos dados que iremos trabalhar.

Assim, propomos uma classificação derivada que melhor se adere à realidade trabalhada:

1. **Fechado** – O dados não estão disponíveis na internet;

2. **Consolidado** – Os dados estão disponíveis somente na forma consolidada, com informações contidas em relatórios em forma de texto;
3. **Abertos** – Os dados estão disponíveis de forma não agregada, em seu grão básico de coleta, em arquivos de formatos proprietários ou livres, ou em sistema de analíticos
 - 3.1. Formatos legíveis apenas por pessoas (pdf);
 - 3.2. Formatos legíveis por computador, em padrão proprietário (xls, mdb);
 - 3.3. Formatos legíveis por computador, em padrão livre (csv, tabwin);
 - 3.4. Ambiente analítico disponível para consultas online;
 - 3.5. Ambiente analítico disponível para consultas online e banco de dados disponível, em padrão livre (csv, tabwin).
4. **Abertos e Ligados** – Os dados estão em formato aberto e ligado (Dado Aberto Ligado).

A seguir, o Quadro 1 apresenta as bases de dados, sua classificação de abertura, os órgãos mantenedores, e os endereços das bases.

Quadro 1: Bases de dados sobre agrotóxicos

Cat.	Sigla - Nome	Abertura	Mantenedor	Link
1	SINAN - Sistema de Informação de Agravos de Notificação	3.5	DataSUS	http://dtr2004.saude.gov.br/sinanweb/
1	SIM - Sistema de Informações sobre Mortalidade	3.5	DataSUS	http://tabnet.datasus.gov.br/cgi/sim/dados/indice.htm
1	SIH-SUS – Sistema de Informações Hospitalares do Sistema Único de Saúde	3.5	DataSUS	http://datasus.saude.gov.br/sistemas-e-aplicativos/hospitalares/sihsus
1	SINITOX – Sistema Nacional de Informações	3.1	Fiocruz	http://www.fiocruz.br/sinitox/

Cat.	Sigla - Nome	Abertura	Mantenedor	Link
	Tóxico-Farmacológicas			
1	CAT - Sistema de Comunicação de Acidente do Trabalho	3.2	MTE	http://www.previdencia.gov.br/estatisticas/
1	NOTIVISA - Sistema Nacional de Notificações para a Vigilância Sanitária	1	ANVISA	http://www.anvisa.gov.br/hotsite/notivisa/relatorios/index.htm
1	Integrador de Registro Hospitalar de Câncer	3.5	INCA	https://irhc.inca.gov.br/
1	Mapa da Injustiça Ambiental e Saúde no Brasil	3.4	Fiocruz	http://www.conflitoambiental.iciet.fiocruz.br
1	PARA - Programa de Análise de Resíduos de Agrotóxicos em Alimentos	2	ANVISA	http://s.anvisa.gov.br/wps/s/r/z
2	SIA - Sistema de Informações Sobre Agrotóxicos	1	ANVISA	http://www4.anvisa.gov.br/agrosia/asp/default.asp
2	AGROFIT - Sistemas de Agrotóxicos Fitossanitários	3.4	MAPA	http://extranet.agricultura.gov.br/agrofit_cons/principal_agrofit_cons
2	LMR - Limite Máximo de Resíduos	3.2	ANVISA	http://s.anvisa.gov.br/wps/s/r/z
3	SIDRA – Censo Agropecuário de 2006	3.5	IBGE	http://www.sidra.ibge.gov.br/bda/pesquisas/ca/
3	SISCOMEX - Sistema Integrado de Comércio Exterior	3.2	Receita Federal	http://www.receita.fazenda.gov.br/Aduana/ProdutosNCM/D

Cat.	Sigla - Nome	Abertura	Mantenedor	Link
				efault.htm
3	SIAGRO - Sistema de Monitoramento do Comércio e Uso de Agrotóxicos do Paraná	1	SEAB/PR	http://www.siaagro.seab.pr.gov.br/siaagro
4	LeXML – Rede de Informação Jurídica e Legislativa	4	GT LexML/Comunidade TIControl	http://www.lexml.gov.br/
4	Banco de Leis da Câmara	3.4	Câmara dos Deputados	http://www.camara.gov.br/sileg/default.asp
5	Agroecologia em Rede	3.4	Articulação Nacional de Agroecologia	http://www.agroecologiaemrede.org.br
6	Portal de Periódicos da CAPES	3.4	CAPES	http://www.periodicos.capes.gov.br
7	Sites de notícias relacionadas à agrotóxicos com RSS	4	Vários	Exemplo: http://pratoslimpos.org.br/?feed=rss2
8	Produção Agrícola Municipal	3.2	IBGE	http://www.sidra.ibge.gov.br/bda/pesquisas/pam/
9	IDH - Índice de Desenvolvimento Humano	3.3	PNUD/ONU	http://www.atlasbrasil.org.br/2013/pt/ranking

Dados abertos ligados

O Quadro 1 apresenta apenas duas bases de dados classificadas na categoria 4. A publicação de mais dados nesta categoria permitiriam a utilização de recursos dentro daquilo que vem sendo chamado de Web Semântica.

A Web tornou-se o principal meio de disseminação de informações digitais, valendo-se especialmente, em um primeiro momento, da publicação e interligação de documentos e, posteriormente, como poderoso canal de comunicação e relacionamento através das redes sociais. Além disso, consolidou-se como *front-end* para acesso a dados, publicados dinamicamente através de aplicações especialmente desenvolvidas para este ambiente. No entanto, muitos desses dados permanecem ainda em silos, onde pouco compartilhamento e interligação de recursos são praticados. Em especial, pode-se considerar a grande quantidade de dados governamentais disponíveis na Web, gerados a partir das muitas iniciativas de *e-gov* e *open-government*, que defendem a ampla divulgação de dados aos cidadãos e organizações. No entanto, o consumo conjunto e a reutilização desses dados ainda são difíceis, dadas suas interfaces voltadas apenas para consulta ou extração *ad-hoc*, além dos altos custos e problemas envolvidos na análise de dados.

Nesse sentido, as iniciativas de dados abertos interligados (Linked Open Data –LOD, como vêm sendo chamadas) propõem o uso de padrões abertos, apoiados pelo W3C, para exposição de dados na Web por meio de princípios simples, inspirados no sucesso da ‘Web de Documentos’, envolvendo padronização da semântica por trás dos dados. Neste sentido, o uso de referências a vocabulários como forma de resolver ambiguidades dos termos utilizados está no centro desta abordagem. LOD emprega tecnologias da Web Semântica no suporte à publicação de dados estruturados na Web e na criação de ligações entre dados de diferentes fontes. Segundo Tim Berners-Lee (2006), seus princípios são:

1. Usar URIs como nomes para recursos;
2. Usar URIs HTTP de forma que pessoas possam procurar por estes nomes;
3. Quando alguém procura uma URI, fornecer informação RDF útil;
4. Incluir sentenças RDF que ligam a outras URIs para que possam descobrir outros recursos.

Nos últimos anos, LOD tem atraído a atenção de órgãos governamentais, onde eco-sistemas de cooperação entre Governo e população podem ser criados através da publicação de dados na Web de uma forma padronizada. Transparência no uso dos recursos públicos e

maximização do retorno sobre dados governamentais públicos, através de colaboração com a população surgem com primeiros casos de uso em dados abertos interligados. Adicionalmente, conforme diversos setores do Governo venham a disponibilizar seus dados como LOD, é esperada maior integração entre diferentes bases de dados, diminuindo as barreiras para a colaboração e para o reúso desses dados em novos contextos dentro do governo. Em síntese, dados abertos interligados potencializam uma melhoria na gestão pública em algumas dimensões, incluindo: (1) maior acessibilidade de dados para os gestores e tomadores de decisão; (2) diminuição da distância entre serviços públicos e usuários, permitindo uma intensificação do diálogo e colaboração entre essas partes; (3) aumento da transparência no uso de recursos públicos.

Dados abertos ligados e o contexto dos agrotóxicos

Para a disponibilização de dados em formato LOD no Portal de Dados Abertos sobre Agrotóxicos, serão utilizados dois vocabulários principais: o Agrovoc e o DataCube. A utilização de vocabulários controlados e ontologias adiciona uma camada semântica aos dados, permitindo seu processamento em conjunto com outras bases de dados que compartilhem os mesmos vocabulários ou o mapeamento entre diferentes vocabulários.

O Agrovoc⁴⁹ é um vocabulário controlado que descreve diversas áreas de interesses ligadas à agência das Organização da Nações Unidas para Alimentação e Agricultura – FAO. As áreas incluem alimentação, nutrição, agricultura, pesca, manejo florestal e meio-ambiente, entre outras. O vocabulário consiste em 32.000 termos em mais de 20 idiomas, incluindo o português.

Como exemplo, podemos citar a relação entre uso de agrotóxicos e a produção de alimentos. O relatório do Programa de Análise de Resíduos em Alimentos (ANVISA, 2010) traz a informação de que foram encontradas doses do agrotóxico Metamidofós em 18 amostras de pimentão. O Censo Agropecuário do IBGE informa a produção de 276.767 toneladas de pimentão em 2006.

No Agrovoc, o termo relativo ao pimentão recebe o código 13370, e tem seu endereço único dado por <<http://aims.fao.org/en/agrovoc-term-info?mytermcode=13370>>. Já o Metamidofós é identificado pelo código

⁴⁹<http://aims.fao.org/standards/agrovoc>

31416 e tem endereço <<http://aims.fao.org/en/agrovoc-term-info?mytermcode=31416>>. Se ambas as bases utilizassem este vocabulário comum, seria possível a terceiros efetuarem análises conjuntas envolvendo os dois tipos de dados.

Já o Data Cube Vocabulary⁵⁰ foi descrito pela W3C, órgão que define padrões na Internet, para anotação de dados estatísticos. Com o Data Cube, é possível descrever em uma base de dados estatísticos conceitos como medidas, dimensões, unidades e bases de dados. Desta forma, é possível desenvolver ferramentas de análise sobre conjuntos de dados estatísticos (também chamados de multidimensionais) utilizando informações sobre as possibilidades e cruzamento e as métricas disponíveis.

Desenvolvimento do Portal

Para o desenvolvimento do portal, foram tomados como base diversos sistemas de dados abertos existentes, como por exemplo o portal de dados abertos do Governo Federal⁵¹, o portal de dados da Cidade do Rio de Janeiro⁵² e o portal de dados abertos do Reino Unido⁵³.

Ferramenta para gestão do conteúdo

Atendendo aos padrões atuais da Web, decidiu-se pela utilização de um Sistema de Gerenciamento de Conteúdo (do inglês Content Management System – CMS). O CMS é uma classe de aplicativos utilizada para gerenciar sítios na Internet de maneira consistente e organizada. O uso de um CMS permite que o conteúdo seja modificado, excluído ou adicionado com facilidade no portal.

O CKAN⁵⁴, desenvolvido pela Open Knowledge Foundation, vem sendo utilizado por diversos governos e grupos de usuário em todo o mundo, e tem impulsionado o uso de portais para publicação de dados

⁵⁰<http://www.w3.org/TR/vocab-data-cube/>

⁵¹<http://dados.gov.br/>

⁵²<http://data.rio.rj.ov.br/>

⁵³<http://data.gov.uk/>

⁵⁴<http://ckan.org/>

abertos. Entre as iniciativas que o utilizam, além dos portais citados acima, está o da União Europeia⁵⁵, entre diversos outros.

O CKAN é uma aplicação Web de código aberto e livre, que fornece um conjunto de ferramentas que permite simplificar a publicação, o compartilhamento e a utilização de dados. Ele pode ser integrado a outros CMS como por exemplo o Joomla ou Drupal, como uma maneira para facilitar o gerenciamento do conteúdo e a publicação dos dados.

O CKAN foi desenvolvido utilizando as linguagens de programação Python e Javascript, PostgreSQL para o banco de dados da aplicação e SQLAlchemy para a camada de abstração entre Python e o banco de dados.

No início do projeto, foi considerado utilizar o CKAN para a implantação do Portal de Dados Abertos sobre Agrotóxicos, iniciando-se o estudo sobre o que seria necessário para fazer uso de suas ferramentas. Durante a fase de testes, foram encontradas algumas dificuldades para a instalação das dependências necessárias para o funcionamento do CKAN no servidor Web.

Os servidores disponíveis para implementação do projeto utilizam o software Apache, que representa uma parcela de mais de 47% dos servidores ativos no mundo, responsável por servir mais de 54% dos sites presentes na Web, de acordo com o site Built With⁵⁶. Buscamos então uma solução para o CKAN, que atendesse nossas necessidades e mantivesse o foco sobre projetos de software livre, e então encontramos o DKAN.

DKAN é uma aplicação web desenvolvida pela NuCivic⁵⁷ para catalogar, publicar e visualizar dados abertos, desenvolvido como uma iniciativa de software aberto e livre, em formato de uma distribuição Drupal. O DKAN é uma iniciativa paralela ao CKAN, desenvolvido para integrar todas as funcionalidades de um portal de dados abertos ao Drupal, com o intuito de facilitar e incentivar a implementação de novas iniciativas de publicação de dados abertos.

O Drupal é um projeto de software livre que segue a licença GPL da GNU, o que significa que o acesso ao código fonte é livre, havendo também a liberdade para criação de extensões. Construído utilizando a linguagem de programação mais usada pelos servidores na Web, o PHP (82,2% de todas as linguagens server-side, contra 17.2% do asp.net⁵⁸), o

⁵⁵<http://publicdata.eu>

⁵⁶<http://trends.builtwith.com>

⁵⁷<http://nucivic.com/>

⁵⁸<http://w3techs.com/>

Drupal separa o conteúdo, armazenado num banco de dados MySQL, da apresentação ao usuário, descrita em linguagens de marcação como HTML, CSS e JavaScript. O núcleo é responsável por todas as funções internas enquanto o usuário final organiza o conteúdo, manipula a aparência, automatiza tarefas administrativas e define permissões e papéis para usuários e colaboradores.

Ainda de acordo com o sítio Built With, o Drupal é o segundo CMS mais utilizado do mundo, cobrindo 0,3% de toda internet (cerca de 783.388 mil sites), ficando atrás apenas do Wordpress (4.56%), que gerencia cerca de 13.164.104 de sítios pela Web. Com relação aos sítios construídos com CMS, o Drupal corresponde a uma parcela de 13%, enquanto o Wordpress é responsável por uma fatia de 50%.

O DKAN pode ser instalado em um servidor Apache com suporte a PHP e MySQL, sem necessidade da instalação de outras dependências, tornando-se muito mais funcional, aliando toda a estrutura de organização de conteúdo de um CMS a uma ferramenta de gestão de conteúdo. Todas as ferramentas disponíveis no CKAN estão presentes no DKAN, entre elas a inclusão de metadados segundo a necessidade do autor, além da visualização dos dados em mapas, tabelas e gráficos.

Visualização dos dados

O Drupal é considerado um *framework* modular, porque permite a adição de módulos que estendem as funcionalidades básicas contidas no núcleo. Entre esses, existe um que é de grande utilidade no Drupal, conhecido como *views*. Este módulo é utilizado para permitir a visualização de *posts*, galerias, tabelas, mapas, gráficos, itens de menu, relatórios, páginas etc. Ele é necessário para permitir a configuração de visualizações no sítio, o que vai aparecer, aonde vai aparecer (controle de contexto) e quem pode ver. Os blocos são seções dentro de uma página e cada bloco é por si uma *view*. O controle de contexto é também um recurso separado das *views* e é nativo do Drupal.

O Drupal possui uma série de recursos e módulos padrão. A criação de tipos de conteúdo é outro módulo nativo, porém requer a instalação de outros módulos quando se deseja criar tipos mais específicos (por exemplo um formulário de contato ou um livro).

O tema usado pelo DKAN requer o uso de 2 módulos essenciais: DKAN *datasets* e DKAN *datastore*. O DKAN *datasets* representa a funcionalidade de gerencia de conjuntos de dados e grupos associados. O

datastore, por sua vez, é o responsável pelo armazenamento, gerencia e visualização de arquivos em vários formatos.

As categorias de dados apresentadas na Seção 2 foram implementadas como grupos do DKAN. Em linhas gerais, os grupos representam uma classificação da(s) área(s) e cada conjunto de dados (*datasets*) se enquadra dentro do domínio. Cada conjunto de dados possui metadados padrão como quem o publicou, data de criação e de modificação e tipo de licença porém também é possível criar atributos personalizados para cada *dataset*.

O módulo *datastore* permite que se associe um ou mais arquivos a um determinado conjunto de dados. Estes arquivos podem ser adicionados de três formas: via um ligação com arquivos remotos, via *upload* do arquivos locais ou via ligação a uma API (*Application Programming Interface* - Interface de Programação de Aplicativos) que retorne dados.

Arquivos em formatos tabulares, como XLS ou CSV, podem ser visualizados e ter suas linhas percorridas *online*, no formato de visualização Grid. O sistema permite ainda visualização em forma de gráficos, após configuração das dimensões, e em formato de mapa, caso sejam especificadas colunas relativas à latitude e longitude.

Conversa sobre os dados

Uma iniciativa de dados abertos certamente não termina com a simples publicação dos dados. É fundamental que as informações estejam disponíveis, mas existem diversos outros fatores determinantes da mobilização que os dados podem gerar. Davies (2012) apresenta uma discussão e uma série de parâmetros para que um projeto de dados abertos possa de fato gerar um engajamento social e ter utilidade política. Estes aspectos são divididos em 5 categorias que representam as 5 estrelas relacionadas aos uso dos dados que um projeto deveria ter (DAVIES, 2012 – tradução livre):

- **Orientado pela Demanda:** As escolhas relacionadas aos dados publicados, sua estrutura e as ferramentas de suporte disponibilizadas são baseadas nas necessidades e demandas da comunidade? Existem meios para ouvir demandas e responder com dados?
- **Contexto:** Existe disponibilização de informações claras para descrever os dados providos, incluindo informações sobre frequência de atualização, formato e qualidade dos dados? Existem informações

qualitativas sobre o conjunto de dados como detalhes de quando foi criado, ou manuais para trabalhar com os dados? Existem ligações com catálogos de dados que sua organização ou outras já mantêm, ou para ferramentas externas e trabalho com os dados?

- **Conversação:** É possível fazer comentários sobre os dados, ou criar conversas estruturadas sobre eles para se conectar com outros usuários dos dados? Você participa das conversas? Existem meios fáceis de se conectar com os mantenedores dos dados na sua organização, para fazer perguntas sobre os dados, ou para que participem das conversas? Existem possibilidades *offline* de conversa sobre os dados?
- **Capacidades, habilidades e redes:** Há a disponibilização de ferramentas ou links para que se possa trabalhar com os dados? Existem Guias de Uso dos Dados para se fazer análises, para que as pessoas possam construir capacidades e habilidades para interpretar e analisar os dados da forma que quiserem? Existe um trabalho direto de formação no uso particular dos dados, ou de dados em geral? Existem formas de incentivo direto ao uso dos dados pela comunidade?
- **Colaboração:** Existem caminhos de retro-alimentação para que o público possa melhorar os conjuntos de dados? Você colabora com a comunidade para a criação de novos conjuntos de dados (dados derivados)? Existe um suporte para a criação de ferramentas e serviços para uso dos dados? Existe um trabalho de colaboração com outras organizações para ligação entre fontes de dados?

Seguindo esta perspectiva, foram analisadas algumas formas de engajamento possíveis no contexto do Portal da Dados Abertos Sobre Agrotóxicos.

Faz-se necessário promover formas de ouvir o que o público, além da comunidade acadêmica, tem a dizer. Gerar uma discussão social contínua em torno do assunto é essencial para a efetivação deste objetivo. Dentro desta linha, o Portal permite que qualquer pessoa envie, sugira um arquivo ou mesmo uma base de dados dentro do domínio.

O Portal precisa funcionar como um grande fórum aberto e permitir comentários, perguntas, respostas de forma dinâmica e descomplicada. Nesse sentido, cada *dataset* ou arquivo a este associado pode receber comentários e gerar discussões. O próximo passo neste sentido é criar formas de interações gerais que possibilitem trocas entre os desenvolvedores e os usuários, e entre os usuários.

Conclusões

Neste artigo, apresentamos o desenvolvimento do Observatório de Atenção Permanente ao Uso de agrotóxicos. O Observatório se concretiza em um portal de dados abertos sobre agrotóxicos, que reúne dados sobre o assunto em diversos eixos. O objetivo do Portal é fornecer um subsídio à sociedade em geral, e em particular aos movimentos sociais, nas estratégias de redução do uso de agrotóxicos e de construção de um modelo de produção de alimentos baseado na agroecologia.

O recente movimento por transparência em diversos países do mundo vem resultando em diversos portais de dados abertos de governos, e também sobre temas específicos. Desenvolveram-se, por consequência, diversas ferramentas de informáticas para facilitar a publicação, o uso e a manutenção das bases de dados.

Entretanto, o desenvolvimento de pesquisas e tecnologias teve maior foco na publicação dos dados. Ao analisarmos os portais do ponto de vista dos usuários, podemos perceber que a publicação é apenas o primeiro passo. É necessário criar as condições para que a sociedade possa se apropriar dos dados de forma concreta, e não ficar somente limitado ao resultado das análises feitas por *experts* em dados.

Nesse sentido, o lançamento da primeira versão do Portal da Dados Abertos Sobre Agrotóxicos inicia um necessário trabalho de organização e divulgação de informações. A partir do lançamento do portal, será então iniciado o trabalho de mobilização para que os dados possam de fato ser utilizados pela sociedade e movimentos sociais, embasando suas reivindicações por uma sociedade em que a saúde e a vida sejam a prioridade na definição do modelo de produção de alimentos.

Bibliografia

ANVISA. **Programa de Análise de Resíduos de Agrotóxicos em Alimentos (PARA)** - Relatório de Atividades de 2010.

BERNERS-LEE, T. **Linked Data - Design Issues**. W3C Website, 2006. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 3 ago. 2014.

BERNERS-LEE, T. **5 - Open Data**. Disponível em: <<http://5stardata.info/>>. Acesso em: 3 ago. 2014.

CARNEIRO, F. F. et al. Dossiê ABRASCO - **Um alerta sobre os impactos dos Agrotóxicos na Saúde, Parte 1 - Agrotóxicos, Segurança Alimentar e Saúde, 2012.** Disponível em <<http://greco.ppgi.ufrj.br/DossieVirtual/>>. Acesso em: 3 ago. 2014.

DAVIES, T. **Supporting open data use through active engagement.** In: Proceedings of the W3C USING OPEN DATA WORKSHOP, Brussels, 2012.