

Informação para Ação: Desenvolvimento de um Portal de Dados Abertos Sobre Agrotóxicos¹

Information for Action: Development of an Open Data Portal about Pesticides

Alan Freihof Tygel²
Leonardo Gonçalves³
Mayara Santos⁴
Gabriel Marques⁵
Maria Luiza Machado Campos⁶

Artigo recebido para publicação em jan./2015 e aceito para publicação em mar./2015

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 tem 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, um portal de dados abertos sobre uso de agrotóxicos foi desenvolvido com objetivo de ser uma ferramenta que possa fornecer subsídios para ações contra o uso de venenos e em defesa da vida. Neste projeto, foi implementado o sistema de gestão de conteúdo DKAN, que permite a disponibilização e visualização de conjuntos de dados através de um portal na Internet, seguindo uma tendência mundial de publicação de informações através de portais de dados abertos.

Palavras-chave: Dados abertos; agrotóxicos; informática e sociedade; saúde; portal de dados abertos.

ABSTRACT

In 2008, Brazil became the world's leader in pesticides consumption, place still held in 2014. The use of more than 1 billion liters of poisons per year is causing great impacts on public health and on the environment. However, the lack of trusty data on various aspects around the use of pesticides and its effects hampers the action of public sector and social movements on denouncing the intoxications and announcing alternatives. In this sense, an open data portal about use of pesticides in Brazil was developed aiming to be an open data portal which could subsidize actions against the use of poisons and in defense of life. In this project, it was implemented a content management system called DKAN, which allows the release and visualization of data sets through an Internet portal, according to an international trend information release through open data portals.

¹O portal foi desenvolvido no âmbito do projeto "Observatório de Atenção Permanente ao Uso de Agrotóxicos", contemplado no Edital ExtPesq da FAPERJ (14/2012) sob o código E-26/111.544/2012

² Doutorando em Informática no PPGI/UFRJ, Engenheiro e Mestre em Engenharia Eletrônica pela UFRJ. Interesse em Dados Abertos Interligados, Políticas de Transparência e Governo Aberto, e Desenvolvimento participativo. Contato: alantygel@ppgi.ufrj.br

³ Graduando em Ciência da Computação na UFRJ. Contato: leonardo.goncalves.2711@gmail.com

⁴ Graduanda em Ciência da Computação na UFRJ. Contato: mayara.morais.santos@gmail.com

⁵ Graduando em Ciência da Computação na UFRJ. Contato: gabrielmarques92@hotmail.com

⁶ Docente no Departamento de Ciência da Computação da UFRJ, e no Programa de Pós-Graduação em Informática da UFRJ. Interesse em Dados Abertos Interligados e Web Semântica. Contato: mluiza@ppgi.ufrj.br

Keywords: *Open Data; Pesticides; Informatics and Society; Health; Open Data Portal.*

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 Vegetal, 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).

De acordo com o Programa de Análise de Resíduo de Agrotóxicos em Alimentos – PARA (ANVISA, 2011), em apenas em 37% das amostras analisadas no país 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 Vida⁷. A campanha visa combater a utilização de agrotóxicos e a ação de empresas associadas a eles. 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 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.

⁷ <http://contraosagrototoxicos.org/>

Assim, este artigo descreve o desenvolvimento de um Portal de Dados Abertos sobre Agrotóxicos⁸, que possa responder a este anseio da comunidade de pesquisadores e movimentos sociais que compõem a Campanha Permanente Contra os Agrotóxicos e Pela Vida.

O restante deste artigo está dividido da seguinte maneira: a seção a seguir apresenta alguns estudos sobre as Tecnologias da Informação e Comunicação (TICs) e suas relações com a sociedade, e mais especificamente a análise sobre os portais de dados abertos. Em seguida, detalhamos aquilo que denominamos dados sobre agrotóxicos, apontando algumas bases de dados e classificando seus níveis de abertura. A sessão posterior discute a tecnologia dos dados abertos interligados, e suas possibilidades no contexto dos agrotóxicos.

Finalmente, descrevemos a implementação do portal, as tecnologias utilizadas e a forma como se espera que ele interaja com a sociedade. Concluímos com algumas indicações para o futuro da ferramenta e seus desafios, que se encontram sobretudo nas dificuldades de apropriação das ferramentas de manipulação de dados pela população em geral.

A SOCIEDADE E AS TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO

As implicações das tecnologias na sociedade, assim como as determinações sociais sobre a construção das tecnologias vêm sendo estudadas pelo campo de conhecimento denominado Estudos Sociais da Tecnologia ou Estudos CTS (Ciência, Tecnologia e Sociedade). Aponta-se para uma não-neutralidade das técnicas, bem como para a necessidade de se desenvolver novas tecnologias para grupos contra-hegemônicos (ALVEAR, 2014, cap. 2).

Em especial, um foco recente vem sendo dado às TICs. Diversos aspectos da relação entre as TICs e a sociedade vêm sendo debatidos, como: os impactos socioculturais das novas TICs (FERN; ALVAREZ, 2011); a relação entre as TICs e a qualidade de vida no envelhecimento (ROCHA et al., 2011); o governo eletrônico/aberto, suas promessas e implicações (DIAS, 2012; VERCELLI, 2012); a relação das TICs com as manifestações de junho 2013 no Brasil (RAMOS; OLIVEIRA, 2014); o uso das TICs pelo movimento social de Economia Solidária

⁸ <http://www.dados.contraosagrototoxicos.org>

(FONSECA; MACHADO, 2013); e as possibilidades do uso das TICs na educação (ARAÚJO; PILLOTTO, 2014).

Neste artigo, chamamos a atenção em especial para uma aplicação das TICs: os portais de dados abertos, desenvolvidos em sua maioria por governos, mas que também podem surgir como iniciativas da sociedade⁹, como será o caso analisado aqui.

PORTAIS DE DADOS BAERTOS, TRANSPARÊNCIA E SOCIEDADE

Recentemente, governos de diversos países vêm lançando portais de dados abertos governamentais, a partir do discurso de aumento da transparência. Com eles, se pretende disponibilizar para a população dados em formato aberto sobre alguns dos temas relacionados à administração pública. Portais de dados abertos podem abrigar desde informações sobre orçamento e gastos públicos, até dados georreferenciados sobre a localização das escolas públicas ou a posição dos ônibus de uma determinada cidade.

O portal de dados abertos dos Estados Unidos da América foi lançado em 2009, com um anúncio já no primeiro dia de governo do presidente Barak Obama. No mesmo ano, o Reino Unido lançou seu portal de dados abertos governamentais. Em ambos os casos, a política veio como uma resposta à sociedade aos pedidos de mais transparência. Hoje, de acordo com o portal estadunidense¹⁰, existem 46 países com portais de dados abertos, e 163 portais de regiões internacionais, incluindo estados, municípios e organismos internacionais como a União Europeia e a Organizações das Nações Unidas (ONU).

Apesar do curto período histórico para análises, alguns autores se desafiaram a entender o real impacto destes portais na sociedade. De acordo com Huijboom e Broek (2011), o objetivo dos portais de dados abertos é aumentar a transparência, a participação e eficiência dos governos. Entretanto, devido ao curto espaço de tempo, ainda é difícil avaliar o impacto destes portais. A conclusão de uma comparação entre 5 portais federais é de que faltam evidências precisas sobre os efeitos

⁹ Neste caso, estamos nos referindo à “sociedade” enquanto movimentos sociais organizados, grupos, ONGs, associações ou mesmo indivíduos que possuem condições (educação, infra-estrutura) para lidar de alguma forma com os dados abertos. Com isso, pretendemos deixar claro que, mesmo usando de forma imprecisa o termo “sociedade”, entendemos que a maior parte da população se encontra excluída deste debate.

¹⁰ <http://data.gov/open-gov/>, acesso em 26/12/2014.

econômicos, sociais e democráticos, e de que é fundamental a obtenção de mais conhecimento para subsidiar os futuros investimentos dos governos.

De acordo com Davies e Edwards (2012), há por trás do discurso dos dados abertos também uma expectativa de transformação social, através da criação de oportunidades para países em desenvolvimento. No trabalho, os autores buscaram revelar, sob uma perspectiva histórica e socio-técnica, o potencial dos dados abertos em desafiar as estruturas de poder na produção e consumo de conhecimento. Segundo os autores, “o desenvolvimento dos dados abertos e interligados se refere tanto a mudanças culturais, organizacionais e normativas quanto à tecnologia”.¹¹

No Brasil, a Instrução Normativa n. 4 instituiu em 2012 a Infraestrutura Nacional de Dados Abertos – INDA (GOVERNO FEDERAL, 2012), cuja atribuição principal foi o desenvolvimento do portal de dados abertos do governo brasileiro¹². Miranda (2011) sistematiza um percurso histórico sobre as iniciativas que culminaram na atual política de dados abertos, lembrando que ainda em 2000, quando da instituição do Governo Eletrônico Brasileiro, dois dos seus princípios se referiam à “promoção da cidadania” e à “inclusão digital”. A mesma publicação onde se encontra este artigo contém outras análises interessantes sobre a política de dados abertos no Brasil, em que pese a grande maioria dos autores seja de funcionários do governo responsáveis pela implementação destas políticas.

A geração sistemática de dados, seja através de pesquisas de campo ou de compilações de sistemas de informação, é uma atividade que requer uma quantidade significativa de recursos, sobretudo em termos de estrutura. Este fato torna o Estado um ente privilegiado quando se trata de publicação de dados.

Entretanto, nos últimos anos, diversas iniciativas oriundas da sociedade civil têm buscado publicar dados que demonstrem visões alternativas em relação àquela mostrada pelo Estado a partir dos seus dados. Podemos classificar estas iniciativas em dois tipos: (i) dados gerados pela sociedade, e (ii) dados do estado vistos pela sociedade. No primeiro caso, movimentos sociais produzem dados para publicação, e no segundo, dados fornecidos pelo governo são utilizados para algum tipo de visão alternativa que revela aspectos até então obscuros. Normalmente, ambos os

¹¹ No original: “The development of open and linked data is as much about organisational, cultural and norm changes as it is about technologies”.

¹² <http://www.dados.gov.br>

casos têm como objetivo alguma visão contestatória. O Quadro 1 mostra alguns exemplos.

Projeto	Endereço	Categoria
Banco de Dados sobre Violência no Campo – Comissão Pastoral da Terra	http://www.cptnacional.org.br/	i
Mapa de Injustiça Ambiental – Rede Brasileira de Justiça Ambiental	http://www.conflitoambiental.icict.fiocruz.br/	i
Mapa de Injustiça Ambiental de MG - Gesta	http://conflitosambientaismg.lcc.ufmg.br/	i
Agroecologia em Rede	http://agroecologia.org.br/	i
Quem são os Proprietários do Brasil?	http://proprietariosdobrasil.org.br/	ii
Donos do Congresso	http://donosdocongresso.com.br/	ii
Olho Neles	http://olhoneles.org/	ii

Quadro 1: Alguns exemplos de portais de divulgação de dados mantidos pela sociedade civil, de acordo com as categorias: (i) dados produzidos pela sociedade e (ii) dados produzidos pelo Estados e vistos pela sociedade. Fonte: autoria própria.

A seguir, apresentamos uma análise sobre os dados sobre agrotóxicos, suas distintas naturezas, as bases de dados disponíveis e seu nível de abertura.

DADOS SOBRE AGROTÓXICOS, UMA DEMANDA DA SOCIEDADE

O acompanhamento das atividades da sociedade organizada em torno Campanha Permanente Contra os Agrotóxicos e Pela Vida mostrou um comportamento aparentemente paradoxal deste movimento em relação ao uso de dados. Por um lado, os dados assumem uma importância fundamental na estratégia de comunicação: os primeiros materiais de divulgação ressaltavam dois dados: “o Brasil é o maior consumidor de agrotóxicos do mundo”, e “cada brasileiro consome 5,2 litros de agrotóxicos por ano” (CAMPANHA PERMANENTE CONTRA OS AGROTÓXICOS E PELA VIDA, 2011). Por outro lado, por motivos diversos, nunca se investiu na produção de mais dados, nem mesmo na reivindicação ao poder público por maior qualidade nas informações.

Neste contexto, foi proposto o desenvolvimento do Observatório de Atenção Permanente ao Uso de Agrotóxicos, como um portal de dados abertos sobre

agrotóxicos. Este portal deve reunir dados que sirvam como subsídio para entender o contexto em que ocorre a utilização dos venenos no país.

Na primeira reflexão coletiva sobre os tipos de dados necessários, chegou-se rapidamente à conclusão de que áreas de conhecimento tão distintas quanto Saúde Coletiva, Química, Biologia, Ecologia, Economia, Direito, Agricultura, Agronomia, dentre outras, devem se juntar para que se possa ter uma visão que alcance o nível de complexidade do problema. O Quadro 2 mostra o resultado de uma “chuva de ideias” entre participantes de uma reunião sobre o tema:

saúde	poluição	lutas	bactericida
meio ambiente	cegueira	vasilhames	acaricida
água	aborto	saúde alternati- va	agricultura fami- liar
solo	câncer	alimentação saudável	imposto
pulverização aérea	veneno	fungicida	empresas
biodiversidade	legislação	populações tra- dicionais	conflito
modelo de agri- cultura	doença	desequilíbrio ambiental	suicídio
agroecologia	alimento	transgênico	agroveneno
legislação	revolução verde	aquecimento global	agricultura cam- ponesa
políticas públi- cas	agricultura con- vencional	sustentabilidade	agricultura orgâ- nica
contaminação	mecanização agrícola	inseticidas	
morte	agroquímicos	soberania ali- mentar e nutri- cional	

Quadro 2: Palavras relacionadas ao termo "agrotóxico", recolhidas durante uma reunião sobre o tema. Fonte: autoria própria.

A partir da análise dos temas 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.

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, trabalho, educação, autonomia 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 (2010). 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 seguinte), e ainda há muito poucos exemplos nesse sentido nos dados que iremos trabalhar.

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

1. **Fechados** – Os dados não estão disponíveis na internet, ou possuem algum tipo de licença que impeça seu uso;

2. **Consolidados** – 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 Interligados** – Os dados estão em formato aberto e interligados (*Linked Open Data*).

A seguir, o Quadro 2 apresenta as bases de dados, sua classificação de abertura, os órgãos mantenedores, e os endereços das bases.

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/hospitala

Cat.	Sigla - Nome	Abertura	Mantenedor	Link
				res/sihsus
1	SINITOX – Sistema Nacional de Informações Tóxico-Farmacológicas	3.1	Fiocruz	http://www.fiocruz.br/sinitox/
1	CAT - Sistema de Comunicação de Acidente do Trabalho	3.2	MTE	http://www.pr.evidencia.gov.br/estatisticas/
1	NOTIVISA - Sistema Nacional de Notificações para a Vigilância Sanitária	1	ANVISA	http://www.anvisa.gov.br/site/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.conflictoambiental.icict.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	Fora do Ar
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/Default.htm
3	SIAGRO - Sistema de	1	SEAB/PR	http://www.sia

Cat.	Sigla - Nome	Abertura	Mantenedor	Link
	Monitoramento do Comércio e Uso de Agrotóxicos do Paraná			gro.seab.pr.gov.br/siagro
4	LeXML – Rede de Informação Jurídica e Legislativa	4	GT Lex-ML/Comunidade de 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	CAPEL	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

Quadro 3: Bases de dados sobre agrotóxicos.

DADOS ABERTOS INTERLIGADOS

O Quadro 3 apresenta apenas duas bases de dados classificadas na categoria 4. A publicação de mais dados nesta categoria permitiria 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 interface de 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 divulgação de alguns 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 Berners-Lee (2006), seus princípios são:

1. Usar URIs (*Uniform Resource Identification*) 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 em formato RDF (*Resource Description Format*) 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 ecossistemas 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 reuso 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 Agrovoc¹³ é um vocabulário controlado que descreve diversas áreas de interesses ligadas à agência da 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, 2011) 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 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.

¹³ <http://aims.fao.org/standards/agrovoc>

Já o Data Cube Vocabulary¹⁴ 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.

A próxima seção apresenta a construção do portal de dados abertos sobre agrotóxicos. O portal tem como objetivo reunir informações sobre os vários aspectos que cercam o tema. A grande maioria dos dados reunidos é produzida pelo Estado, mas também serão disponibilizadas algumas bases produzidas pelos movimentos sociais.

DESENVOLVIMENTO DO PORTAL

Para o desenvolvimento do portal, foram tomados como base diversos sistemas de dados abertos existentes, por exemplo, como o portal de dados abertos do Governo Federal, o portal de dados da Cidade do Rio de Janeiro¹⁵ e o portal de dados abertos do Reino Unido¹⁶.

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 CKAN¹⁷, 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 abertos. Entre as iniciativas

¹⁴ <http://www.w3.org/TR/vocab-data-cube/>

¹⁵ <http://data.rio.rj.ov.br/>

¹⁶ <http://data.gov.uk/>

¹⁷ <http://ckan.org/>

que o utilizam, além dos portais citados acima, está o da União Europeia¹⁸, 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 sistema 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 With¹⁹. 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.

O DKAN é uma aplicação web desenvolvida pela NuCivic²⁰ 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,

¹⁸<http://publicdata.eu>

¹⁹<http://trends.builtwith.com>

²⁰<http://nucivic.com/>

contra 17.2% do asp.net)²¹, o 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. O sistema possui ainda uma comunidade extremamente ativa.

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, denominado *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.

²¹ <http://w3techs.com/>

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: *Datasets* e *Datastore*. O *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 das áreas, 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 uma 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 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 retroalimentação 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 de 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 visitantes sugiram bases de dados, que posteriormente serão analisadas pela equipe de manutenção, e publicadas, caso seja pertinente.

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 no Brasil. 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 informática 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 o povo 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. No momento do envio deste artigo, o portal está sendo alimentado com

bases de dados e pode ser visto no endereço <<http://www.dados.contraosagrototoxicos.org/>>.

Consideramos que o simples acesso à informação não é suficiente para a transformação de uma realidade tão consolidada quanto é o uso de agrotóxicos no Brasil. Entretanto, consideramos que ele é uma condição necessária para alimentar a organização e mobilização da sociedade em favor de uma transformação completa do modelo de produção agrícola, do agronegócio produtor de *commodities* para exportação, baseado em agrotóxicos e transgênicos, para uma agricultura familiar camponesa produtora de alimentos saudáveis, baseada na agroecologia.

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.

REFERÊNCIAS

ALVEAR, C. A. DE S. Tecnologia e participação: sistemas de informação e a construção de propostas coletivas para movimentos sociais e processos de desenvolvimento local. **Tese de Doutorado. Programa de Engenharia de Produção, Coppe, Universidade Federal do Rio de Janeiro, 2014.**

ANVISA. Programa de Análise de Resíduos de Agrotóxicos em Alimentos (PARA) - Relatório de Atividades de 2010. **Brasília: 2011.**

ARAÚJO, P. K. H. DE; PILLOTTO, S. S. D. Tecnologia e Educação: O blog como possibilidade de mediação cultural e novas aprendizagens. **Revista Tecnologia e Sociedade**, n. 19, p. 99 – 143, 2014.

BERNERS-LEE, T. Linked Data - Design Issues. **W3C Website**, 2006.

BERNERS-LEE, T. **5 ★ Open Data**. Disponível em: <<http://5stardata.info/>>. Acesso em: 16 jan. 2014.

CAMPANHA PERMANENTE CONTRA OS AGROTÓXICOS E PELA VIDA. **Cartaz n.1** Brasília, 2011. Disponível em: <<http://www.contraosagrototoxicos.org/index.php/materiais/cartazes/cartaz-da-campanha-2011-a4/download>>. Acesso em 25/02/2015.

CARNEIRO, F. F. et al. **Dossiê ABRASCO - Um alerta sobre os impactos dos Agrotóxicos na Saúde, Parte 1 - Agrotóxicos, Segurança Alimentar e Saúde**. Rio de Janeiro: 2012.

DAVIES, T. **Supporting open data use through active engagement** Anais do W3C Using Open Data Workshop, Bruxelas, 2012

Revista Tecnologia e Sociedade, Curitiba, v. 11, n. 22, 2015
Edição Especial XII Encontro Nacional de Engenharia e Desenvolvimento Social
ISSN (versão online): 1984-3526
ISSN (versão impressa): 1809-0044

DAVIES, T.; EDWARDS, D. Emerging Implications of Open and Linked Data for Knowledge Sharing in Development. **IDS Bulletin**, v. 43, p. 117–127, 2012.

DIAS, R. D. B. Governo eletrônico: ferramenta democrática ou instrumento do neoliberalismo? **Revista Tecnologia e Sociedade**, n. 15, p. 30–44, 2012.

FERN, A.; ALVAREZ, L. Algunas consideraciones sobre el impacto sociocultural de las nuevas tecnologías de la información y las comunicaciones. **Revista Tecnologia e Sociedade**, v. 1, n. 13, 2011.

FONSECA, R. DA S.; MACHADO, L. R. DE S. Indicadores de Penetração e Uso da Internet por Empreendimentos da Economia Solidária. **Revista Tecnologia e Sociedade**, n. 16, p. 54–64, 2013.

GOVERNO FEDERAL. **Infraestrutura Nacional de Dados Abertos – INDA**. . 2012, p. INSTRUÇÃO NORMATIVA No 4.

HUIJBOOM, N.; BROEK, T. VAN DEN. Open data: an international comparison of strategies. **European Journal of ePractice**, n. 12, p. 1–13, 2011.

MIRANDA, C. M. C. **A Disseminação de Dados Governamentais como Serviço Público – Os Dados Abertos Governamentais e a Experiência Brasileira** Dados Abertos para a Democracia na Era Digital. **Anais...**Brasília: Fundação Alexandre Gusmão, 2011

RAMOS, A.; OLIVEIRA, R. Indivíduos, sociedade, tecnologia: as manifestações nas ruas das cidades brasileiras e as redes sociais. **Revista Tecnologia e Sociedade**, n. 20, p. 176–187, 2014.

ROCHA, J. D. P. et al. Tecnologias de informação e comunicação, envelhecimento humano e qualidade de vida em diferentes contextos sociais. **Revista Tecnologia e Sociedade**, n. 13, 2011.

VERCELLI, A. Gobiernos y ciudadanía abiertas en la era digital: análisis socio-técnico sobre la participación política y la gestión de bienes intelectuales. **Revista Tecnologia e Sociedade**, n. 15, p. 7–29, 2012.