

UMA ABORDAGEM PARA CAPTURA DE INFORMAÇÕES SOBRE APLICAÇÕES QUE FAZEM USO DE DADOS ABERTOS

RESUMO

As recentes iniciativas de abertura de dados, permitindo que qualquer pessoa possa livremente usá-los, reutilizá-los e redistribuí-los, somado ao avanço da interação entre governo e sociedade através da Internet, têm contribuído para que todo cidadão tenha acesso à informação pública. Além disso, quando compartilhado abertamente os dados têm seu valor e seu uso potencializados. Existem vários portais que permitem a catalogação de dados abertos, como o dados.gov.br que é o Portal Brasileiro de Dados abertos. Dentre as plataformas disponíveis para criação de portais de dados, destaca-se o software open source CKAN, que é bastante difundido no domínio de dados abertos e está presente nos principais portais governamentais do mundo inteiro. Esses catálogos de dados são descritos através de vocabulários, que basicamente definem os conceitos e classificam os termos, como se fossem verdadeiras ferramentas conceituais. Contudo, ainda é um grande desafio rastrear o uso dos conjuntos de dados e as aplicações que fazem uso desses dados. Dessa forma, o objetivo deste trabalho é desenvolver uma abordagem para captura de informações sobre aplicações que fazem uso de dados abertos. Assim, como parte da abordagem proposta neste trabalho foram desenvolvidos: i) um vocabulário, denominado APP-OD, que contempla os principais termos usados na descrição de aplicações que fazem o uso de dados abertos, ii) um repositório online, denominado APPCKan, para cadastro, visualização e busca das descrições dessas aplicações e iii) uma extensão para a ferramenta CKAN, que permite a ligação do portal de dados abertos com o repositório APPCKan.

PALAVRAS-CHAVES: Vocabulário RDF; CKAN; APP-OD; AppCKAN; Open Data.

AN APPROACH FOR CAPTURING INFORMATION ABOUT APPLICATIONS THAT MAKE USE OF OPEN DATA

ABSTRACT

The recent initiatives to open data, allowing anyone to use, to reuse and to redistribute them freely, plus the advancement of interaction between government and society through the Internet, have contributed to that every citizen can have access to public information. Moreover, data has value and its use are intensified when shared openly. There are several portals that allow the cataloguing of open data, such as the Brazilian Portal of Open Data (dados.gov.br). Among the platforms available for creating data portal, stands out the CKAN an open source software, that is widespread in the field of open data and is present in major government portals worldwide. These data catalogues are described by vocabularies, that basically defines the concepts and classify terms as if they were true conceptual tools. However, it is still a real challenge to track usage of the data sets and applications that make use of such data. Thus, this study aims to develop an approach to capturing information about applications that make use of open data. So, as part of the approach proposed in this paper were developed: i) a vocabulary called APP-OD, which includes the main terms used in the description of applications that make use of open data, ii) an online repository called APPCKan for registration, visualization, and search of the descriptions of these applications, and iii) an extension for CKAN tool, which allows the connection between the open data portal and APPCKan repository.

KEYWORDS: RDF Vocabulary; CKAN; APP-OD; AppCKAN; Open Data.

*Revista Brasileira de
Administração Científica,
Aquidabã, v.5, n.2, Out 2014.*

ISSN 2179-684X

SECTION: Articles

TOPIC: *Sistemas e Tecnologia da
Informação*



*Anais do Simpósio Brasileiro de
Tecnologia da Informação (SBTI 2014)*



DOI: 10.6008/SPC2179-684X.2014.002.0010

**Lairson Emanuel Rodrigues
de Alencar Oliveira**

Universidade Federal de Pernambuco, Brasil
<http://lattes.cnpq.br/1146119227551894>
lerao@cin.ufpe.br

Bernadette Farias Lóscio

Universidade Federal de Pernambuco, Brasil
<http://lattes.cnpq.br/2512064355660153>
bfl@cin.ufpe.br

Received: 31/08/2014

Approved: 15/10/2014

Reviewed anonymously in the process of blind peer.

Referencing this:

OLIVEIRA, L. E. R. A.; LÓSCIO, B. F.. Uma abordagem para captura de informações sobre aplicações que fazem uso de dados abertos. *Revista Brasileira de Administração Científica*, Aquidabã, v.5, n.2, p.127-140, 2014. DOI: <http://dx.doi.org/10.6008/SPC2179-684X.2014.002.0010>

INTRODUÇÃO

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 (FONSECA FILHO, 2011). Com a aplicação de técnicas e ferramentas adequadas, torna-se possível gerar novos conhecimentos, informações úteis e melhores formas de visualização dos dados públicos governamentais (MOREIRA, 2013).

A Open Knowledge Foundation (OKF), define que dados são abertos quando qualquer pessoa pode livremente usá-los, reutilizá-los e redistribuí-los, estando sujeito, no máximo, a exigência de creditar a sua autoria e compartilhar pela mesma licença (OKF, 2007). A Cartilha Técnica para Publicação de Dados Abertos no Brasil (BRASIL, 2012) acrescenta 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 no Brasil em 16 de maio de 2012, sendo permitido a qualquer pessoa ter acesso a documentos e informações que estejam sob a guarda de órgãos públicos, em todos os poderes e níveis do governo. Em dezembro de 2011, foi lançado o Portal Brasileiro de Dados Abertos, o dados.gov.br, com o objetivo de promover a interlocução entre atores da sociedade e o governo. Tal como 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. Para isso, utiliza o software CKAN, que além de gerenciar a publicação, uso e compartilhamento dos dados, é open source, extensível e personalizável (CKAN, 2014).

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 (W3C, 2012).

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. Assim, um vocabulário voltado para a descrição do uso dos dados e a implementação de uma ferramenta que capture essas informações torna-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 (W3C, 2012).

Dessa forma, o objetivo deste trabalho é desenvolver uma abordagem para captura de informações sobre o uso de dados abertos. Como parte da abordagem proposta, foram desenvolvidos um vocabulário para descrição de informações relevantes sobre o uso de dados, um repositório online para cadastro de aplicações e uma extensão para a ferramenta CKAN, que liga o portal ao repositório para captura de tais informações.

REVISÃO TEÓRICA

Dados Abertos

A Open Knowledge Foundation (OKF), uma organização da sociedade civil, apartidária e sem fins lucrativos, é uma das maiores incentivadoras do movimento de dados abertos. Ela define que os dados são abertos quando qualquer pessoa pode livremente usá-los, reutilizá-los e redistribuí-los, estando sujeito, no máximo, a exigência de creditar a sua autoria e compartilhar pela mesma licença (OKF, 2007).

Em 2007, um grupo com 30 defensores de dados abertos reuniram-se na Califórnia (EUA) para desenvolver um conjunto de princípios de dados abertos governamentais. Eles chegaram a um consenso e definiram oito princípios para a abertura dos dados do governo. Assim, os dados do governo são considerados abertos se eles se tornarem públicos de maneira que respeitem os princípios abaixo: (1) Completos: todos os dados públicos devem ser disponibilizados; (2) Primários: são apresentados conforme colhidos da fonte; (3) Atuais: devem ser disponibilizados o mais rápido possível para preservar seu valor; (4) Acessíveis: são disponibilizados para a maior quantidade possível de pessoas; (5) Compreensíveis por máquina: devem ser estruturados de modo razoável, para permitir que sejam processados automaticamente; (6) Não discriminatórios: devem estar disponíveis para qualquer pessoa, sem necessidade de cadastros; (7) Não proprietário: nenhuma entidade ou organização deve ter o controle exclusivo sobre os dados disponibilizados; (8) Livres de licenças: Não devem estar submetidos a direitos autorais, patentes, marcas registradas ou regulações de segredo industrial.

Somado a eles, o grupo afirmou que a conformidade precisa ser verificável e uma pessoa deve ser designada como contato responsável pelos dados. Exceto pelo primeiro princípio, que trata de dados do poder público, os demais podem ser aplicados a dados abertos de modo geral. Em 2009, o especialista em políticas públicas e ativista dos dados abertos, David Eaves, propôs três leis para os dados abertos governamentais, as quais também se aplicam a dados abertos de forma geral:

Se o dado não pode ser encontrado e indexado na Web, ele não existe; Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; (3) Se algum dispositivo legal não permitir sua replicação, ele não é útil (EAVES, 2009).

Assim, os dados abertos são pautados pelas três leis e os oito princípios destacados anteriormente. Portanto, devem ser publicados e disseminados na Web, em formato bruto e sem restrições de uso, possibilitando a compreensão automatizada e a criação de aplicações. Geralmente são disponibilizados em JSON, XML, CSV e RDF, que são os principais formatos. (OPEN DATA HANDBOOK, 2010). Este último é o formato recomendado pelo W3C, pela facilidade de combinar dados neste formato com múltiplas fontes de dados, interligando-se a outras iniciativas Open Data na Web.

No governo, surgiu com uma demanda social para aumentar a transparência, a colaboração e a participação dos cidadãos nas políticas e ações (TCE-RS, 2013). Dessa forma, os dados podem ser manipulados a fim de produzir cruzamentos de dados, novas interpretações e novas aplicações (DUTRA & LOPES, 2014), com a possibilidade de uso de diferentes conjuntos de dados para construir sistemas grandes e complexos (W3C, 2011).

Portais de Catalogação de Dados

O ciclo de abertura dos dados, seja no governo ou não, é amplo e precisa ser observado para que as iniciativas de publicação e abertura sejam sustentáveis. Assim, uma das partes chave nesse processo é a de publicação dos dados, que mantém a interface de quem disponibiliza com as pessoas que vão utilizá-los. Depois de ter os dados, se a intenção é publicar ou republicar, é necessário que sejam armazenados em repositórios estruturados e planejados para receber e distribuir dados de forma aberta e interoperável (CÓRDOVA, 2013). Isto é, os dados devem ser disponibilizados nos chamados catálogos de dados, os quais são definidos pelo W3C como sendo uma coleção com curadoria de metadados sobre conjuntos de dados (W3C, 2014).

Um catálogo de dados abertos também pode ser entendido como um mecanismo, ferramenta ou serviço que é responsável pela gestão e publicação de dados e metadados na Web. Assim, como funcionalidade básica, um catálogo deve permitir a busca, inclusão e exclusão de dados e metadados (MIRANDA, 2013). No Brasil, a principal referência na disponibilização dos conjuntos de dados (datasets) é o Portal Brasileiro de Dados Abertos, que é a ferramenta disponibilizada pelo governo para que todos possam encontrar e utilizar os dados e as informações públicas. Ele utiliza o software CKAN, que é uma das ferramentas mais utilizadas nesse contexto.

O CKAN é um software livre e provê um portal de dados, que permite a exposição de catálogos de dados, bem como funções para publicação, armazenamento e gerenciamento dos conjuntos de dados. Ele foi desenvolvido pela OKF e é mantido, hoje, pela comunidade CKAN. Por ser livre, não tem custos e tem uma curva de aprendizado muito positiva (CÓRDOVA, 2013). Conta ainda com uma API para acesso automatizado, pré-visualização de dados, gráficos e mapas, somado à busca de datasets geolocalizada. Além do portal brasileiro, ele também é usado no portal do Uruguai, Canadá, Reino Unido, Estados Unidos, Irlanda, Romênia, Austrália, Holanda, Itália, Argentina e Alemanha.

O Junar Open Data Platform também é outro software usado para portais de catalogação de dados. Ele faz uso do conceito de Software como Serviço (SaaS), tem sua plataforma baseada em nuvem e permite a visualização dos dados em qualquer dispositivo. É usado no portal oficial de dados abertos do Chile e da Costa Rica, bem como no portal de dados abertos da América Latina e algumas cidades nos Estados Unidos e no Mundo (JUNAR, 2014).

Além deles, também existe o Socrata, um software também baseado em nuvem e na ideia de Software como Serviço, seu principal diferencial é na construção de visualizações mais elaboradas, com formatação condicional, gráficos e mapas (MIRANDA, 2013). Ele é usado no portal oficial do Kenya, no portal do Banco Mundial, em cidades e estados dos EUA como Chicago, Nova York, Austin, Maryland e Illinois.

Apesar de ser mais recente, o OpenDataSoft, lançado em 2011, também pode ser encontrado em alguns portais de cidades como Brussels e Paris, além da Córsega, Île de France e o jornal L'Équipe. Seu principal diferencial é a facilidade para buscar, navegar e realizar filtros dentro dos datasets, que geram automaticamente o código para consumo via API.

Desses softwares, podemos destacar o CKAN por ser o único open source, enquanto os outros são pagos. Esse tipo de software tem impacto significativo na redução de custos e diminuição da duplicação de esforços, além da possibilidade de customizações a partir do código fonte e do fato que à medida que o número de desenvolvedores aumenta, maior é possibilidade de identificar e corrigir falhas em menos tempo (GHOSH, 2006).

Aplicativos que fazem uso de dados abertos

A publicação de dados abertos possibilita à sociedade utilizar e reutilizar livremente os dados de acordo com seu interesse e conveniência. O grande potencial desses dados no setor público e privado é amplamente reconhecido, assim como, o impacto que podem causar quanto à transparência, inovação e a economia (LOUTAS et al., 2012). De uma maneira bem geral, podemos dizer que os aplicativos são úteis para gerar visualizações e permitir a análise de dados.

As organizações públicas e privadas têm incentivado e organizado movimentos, conhecidos como hackathons, para estimular a criação de novos aplicativos (SILVEIRA, 2013). São também são chamados de hackday, hackfest e codefest. Neles, programadores, desenvolvedores de software, designers, ativistas, analistas e profissionais de diversas áreas trabalham juntos, utilizando sua inteligência e habilidade para criar conjuntamente soluções tecnológicas, que agregam valor aos dados e, geralmente, os melhores aplicativos são premiados (OLIVEIRA et al., 2013). Conforme evidenciado por Oram (2011), dois grandes problemas são conhecidos ao se realizar essas competições: incentivar os desenvolvedores a transformar seus aplicativos interessantes em produtos sustentáveis e ter público para usá-los. Assim, muitos aplicativos produzidos são rapidamente abandonados.

Os portais de dados, além de disponibilizar os conjuntos de dados, também são utilizados para catalogar os aplicativos que foram criados a partir de seus datasets, incluindo os desenvolvidos a partir de seus concursos. Porém, ao gerar uma amostra com 426 portais de catalogação de dados abertos de 70 países distintos, constatou-se que apenas 28% listam ou tem uma seção para aplicativos e apenas 4% permitem o cadastro, conforme mostrado na Figura 1. A amostra foi gerada coletando os portais presentes no DataCatalogs.org, os disponibilizados nos sites do Socrata, Junar e CKAN, além dos encontrados com o auxílio de um WebCrawle ao consultar pelo termo “open data”. Após a coleta, todos foram acessados e vistoriados considerando os seguintes critérios: se usam ou não software de catalogação de dados, se tem seção que liste os aplicativos desenvolvidos, se permitem cadastro de aplicativos e, permitindo cadastro, se tem integração automática com os seus datasets.

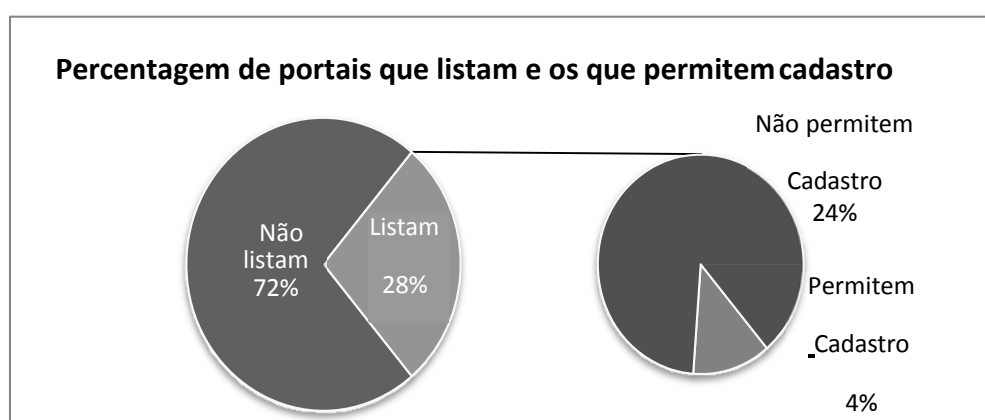


Gráfico 1: Percentagem de portais que listam e os que permitem o cadastro de aplicativos

As aplicações que usam dados abertos no Brasil e no mundo, em sua grande maioria, buscam oferecer meios para melhorar a qualidade de vida dos cidadãos. Através delas e dos dados que as originam, discussões e grupos de estudos interdisciplinares que discutem ideias, analisam os dados e propõem soluções, podem incentivar a sociedade a participar das discussões em torno de assuntos que interessam a todos (OLIVEIRA et al., 2013). Assim, deixando o cenário propício para que novas comunidades sejam formadas e que sejam desenvolvidos aplicativos mais sustentáveis (JOHNSON, 2010). Além disso, se as entidades e os governos contextualizarem de forma mais ampla os dados, com orientações sobre seu significado e/ou relevância, é uma possível estratégia para que os aplicativos se tornem mais efetivos e utilizados (ORAM, 2011).

Encontrar um app que vai resolver determinadas necessidades de um indivíduo não é uma tarefa trivial. Dificuldades com dados desatualizados, descontinuidade do projeto e falta de manutenção, por exemplo, são alguns dos motivos que levam um aplicativo a ficar obsoleto [29]. Soma-se, ainda, o fato que eles geralmente são publicados em vários sites distintos e não são encontrados nos principais portais de dados abertos, conforme evidenciado na Figura 1.

A partir da análise dos dados, constatou-se também que 78% de todos os portais analisados, bem como 83% dos portais que listam ou tem uma seção dedicada para aplicativos, pertencem a

18 países e, dentre eles, os Estados Unidos, Canadá, Itália, França, Reino Unido e Espanha, tem maior representação. Além disso, foram encontrados 17 de portais com formulário para cadastro de aplicações. Mas, apenas cinco deles apresentaram integração com os seus datasets. Vale salientar que não foram considerados os portais que dispuseram apenas de formulários de contato (nome, e-mail e mensagem) para publicar aplicações.

Considerando o gráfico por tipo de software, apresentado na Figura 2, temos que o CKAN foi utilizado por 27% dos portais analisados. Vale salientar que na maior parte da amostra foram detectadas páginas estáticas contendo os conjuntos de dados, ferramentas desenvolvidas pelos próprios governos ou por empresas que não foram consideradas neste trabalho. Ao considerar apenas os 27% dos portais que utilizaram o CKAN, constatou-se que 70% dos portais não listam e não permitem cadastro de aplicações.

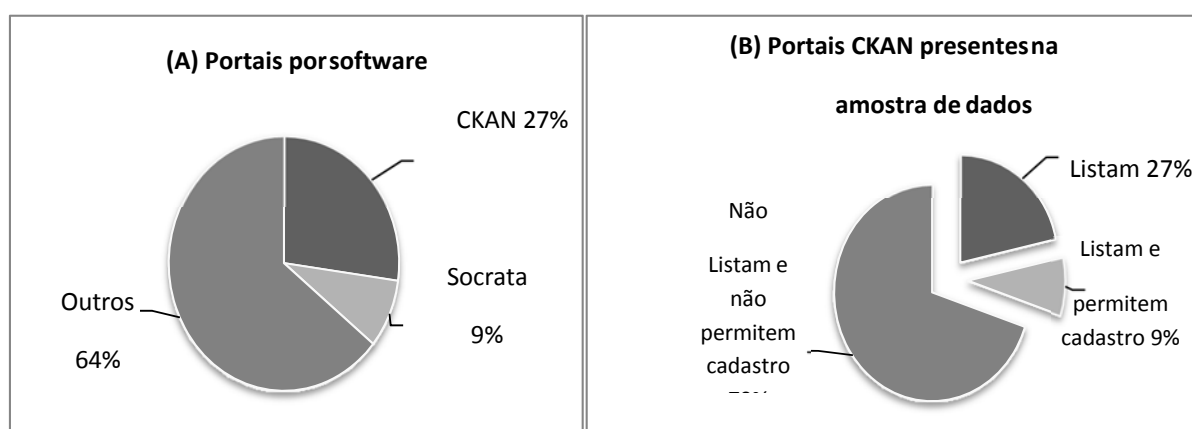


Gráfico 2: (A) Percentagem por software e (B) situação atual considerando apenas o CKAN.

Por fim, ao explorar esses portais, ficou evidente que um repositório central para cadastro de aplicativos desenvolvidos a partir deles, pode facilitar a busca por aplicativos de interesse dos usuários. O uso de um repositório desta natureza pode ajudar na formação de comunidades, que poderão discutir novas ideias para necessidades locais, por exemplo, a partir de uma variedade de aplicativos de diferentes localidades. Além disso, torna-se uma fonte importante de coleta e retorno de informações valiosas, como às relacionadas aos datasets mais usados, os tipos, categorias mais populares e o ciclo de vida das aplicações. Neste contexto, este trabalho propõe desenvolver uma extensão para o CKAN, integrada a um repositório online, para permitir o cadastro dos aplicativos descritos a partir do vocabulário APP-OD.

Visão geral dos vocabulários relacionados

Segundo o W3C, na Web Semântica, os vocabulários definem os conceitos e os termos utilizados para descrever e representar uma área de conhecimento. Também caracterizam possíveis relações e restrições sobre o uso desses termos (W3C, 2013). Assim, nos ajudam a descrever os dados e seus formatos, como se fossem ferramentas conceituais, e possibilitam a descoberta de novas relações entre os dados (W3C, 2013).

Para definir um vocabulário o W3C recomenda o uso de RDFS (RDF Schema) e OWL (Web Ontology Language). Essas linguagens são uma extensão do RDF (Resource Description Framework), que é um modelo de dados para descrição de recursos da Web, fornecendo uma estrutura comum para expressar as informações para serem trocadas entre aplicações sem perda de sentido (W3C, 2014). Assim, essa capacidade de troca de informações significa que a informação pode ser disponibilizada para fins diferentes daqueles para os quais originalmente foram criados, permitindo inclusive a interligação de recursos na Web, como proposto pelos princípios de Linked Data (BERNERS-LEE, 2006).

Segundo Cunha et al. (2011), mesmo não tendo restrições para seleção dos vocabulários, é considerada uma boa prática o reuso de termos de vocabulários RDF amplamente conhecidos para facilitar o processamento de dados pelas aplicações clientes, assim como, novos termos só devem ser definidos se não forem encontrados em vocabulários já existentes. Dessa forma, podemos fazer uso dos vários vocabulários RDF que são utilizados para expressar os dados em formato aberto e que estão sendo largamente utilizados. Dentre eles, podemos destacar o Data Catalog Vocabulary (DCAT), que foi projetado para facilitar a interoperabilidade entre catálogos de dados na Web e o FOAF (o Friend Of A Friend), utilizado para descrever dados pessoais e os relacionamentos interpessoais.

No domínio específico de descrição de aplicações que fazem uso de dados abertos, até o presente momento, não foram encontrados vocabulários recomendados pelo W3C. Todavia, analisando a literatura, foi possível encontrar dois vocabulários descritos em RDF: o odapps¹², que visa facilitar a descrição de ideias e aplicativos baseados em dados abertos, e o DVIA¹³, facilitar a interoperabilidade entre aplicações publicadas na Web. Esses vocabulários podem ser acessados através do portal Linked Open Vocabularies (LOV), mantido pela OKF.

Especificação do vocabulário AP-OD

O primeiro passo para a criação do vocabulário APP-OD foi avaliar quais conceitos, dados e metadados referentes ao cadastro de aplicações são importantes para modelagem coerente do domínio. Foram analisados todos os portais¹⁵ em que foram encontrados formulários para cadastro de aplicações, como o Portal Brasileiro de Dados Abertos, Portal do Reino Unido, Alemanha, Itália, Canadá e Áustria. A análise se deu pela verificação dos campos/dados que cada formulário solicitava, sendo o resultado sumarizado e descrito no Quadro 1.

Foi possível observar que, até mesmo nos formulários mais simples, os seguintes campos/termos eram solicitados: nome do aplicativo, descrição, URL, nome e e-mail do desenvolvedor. Nos formulários mais trabalhados, encontramos a adição de campos extremamente importantes para este trabalho, como categoria, palavras-chaves e, principalmente, os datasets utilizados. Assim, todos foram selecionados para a composição do APP-OD.

Quadro 1: Dados necessários para descrever os aplicativos de dados abertos.

Campo/Termo	Descrição
Nome	Nome do aplicativo
Descrição/Resumo	Breve descrição do aplicativo
Categoria	Categoria do aplicativo
Tags/Keywords	Termos que o descrevem
Tipo de Aplicação	Se o aplicativo é Mobile ou Web
Sistema Operacional	Em qual sistema operacional o aplicativo pode ser executado (ex: Android, BlackBerry, Web, iOS, Windows Phone, entre outros)
Logo	Logo do aplicativo
URL	URL para página Web ou download do aplicativo
Organização	Organização que desenvolveu o aplicativo, se existir
Nome do Desenvolvedor	Pessoa que desenvolveu o aplicativo, se não for organização
E-mail	E-mail de quem desenvolveu o aplicativo
Natureza do Aplicativo	Se o aplicativo é Livre, Parcialmente Livre, Pago
Datasets utilizados	Todos os <i>datasets</i> utilizados
Comentários Adicionais	Comentários adicionais que se façam necessários sobre o aplicativo e/ou desenvolvimento.

Como a proposta, além do vocabulário, é a criação de um repositório central para armazenar descrições sobre as aplicações que fazem uso de dados de múltiplos portais que usam o CKAN, tornou-se necessário acrescentar alguns campos, os quais são descritos no Quadro 2.

Quadro 2: Dados adicionais para descrever os aplicativos de dados abertos.

Campo/Termos	Descrição
Portal de dados	Link do portal CKAN de origem, ou seja, o portal de onde os <i>datasets</i> foram utilizados
Data de cadastro	A data de cadastro da aplicação
Código Fonte	Código fonte do aplicativo, se ele for livre

Consultando tanto os vocabulários de caráter geral como os voltados para a descrição de aplicações, um total de 5 classes, 2 subclasses e 15 propriedades foram reusadas (prefixados com o *namespace* do respectivo vocabulário). Os termos não encontrados foram criados para complementar este novo vocabulário, que recebeu como *namespace* o prefixo APP-OD. A Figura 3 apresenta o diagrama com todas as classes e propriedades do vocabulário APP-OD.

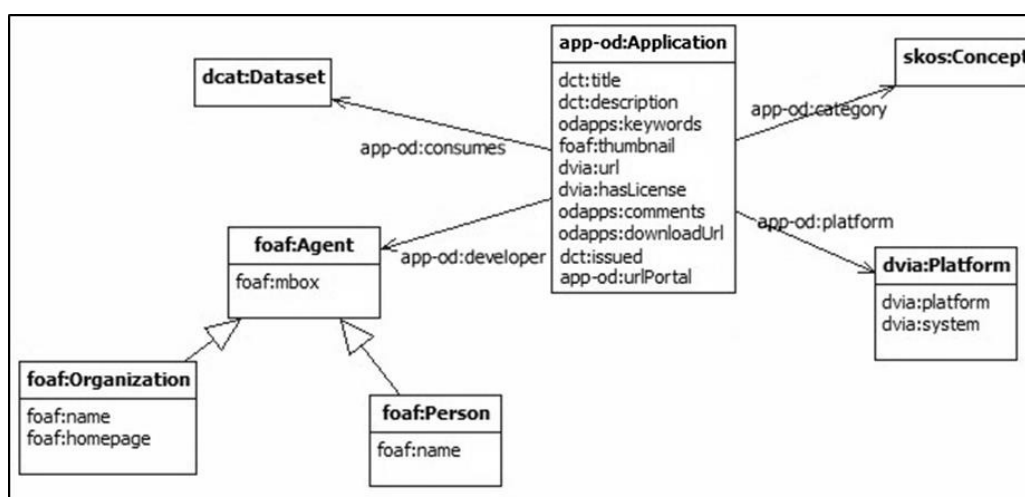


Figura 3: Visão geral do vocabulário APP-OD

A especificação completa deste vocabulário foi publicada no endereço <http://appckan.com/app-od>. Como afirmado anteriormente, os vocabulários ajudam a descrever os dados e seus formatos, como se fossem ferramentas conceituais. Assim também é o objetivo do

APP-OD, que visa facilitar a descrição de aplicativos que fazem uso dos dados abertos e agregar valor a um domínio que, até então, possui poucos vocabulários descritos na literatura.

Extensão para o CKAN

O CKAN foi desenvolvido em Python e faz uso de outras tecnologias e linguagens, como HTML, CSS, Javascript, PostgreSQL e SQLAlchemy. Ele foi projeto para aceitar extensões que permitem estender suas funcionalidades ou modificar as já existentes. Basicamente, uma extensão CKAN é um pacote Python que modifica ou adiciona algo a ele. Cada extensão contém um ou mais *plugins* que devem ser adicionados, posteriormente, ao arquivo de configuração do CKAN para serem ativados. Além disso, é composta por um arquivo chamado *setup.py*, que é o script de instalação, e outro denominado *plugin.py*, que contém os recursos (CKAN, 2013).

O objetivo da extensão desenvolvida neste trabalho é obter a URL do portal CKAN e adicionar em seu *template* um botão com o link para a aplicação online APPCKan, enviando como parâmetro sua URL. Na Figura 3 é apresentada a classe que acrescenta o botão no *template* dos portais e o resultado final. Todos os códigos para instalação foram disponibilizados no endereço <https://github.com/lerao/ckanext-appckan>, além dos procedimentos para instalação.

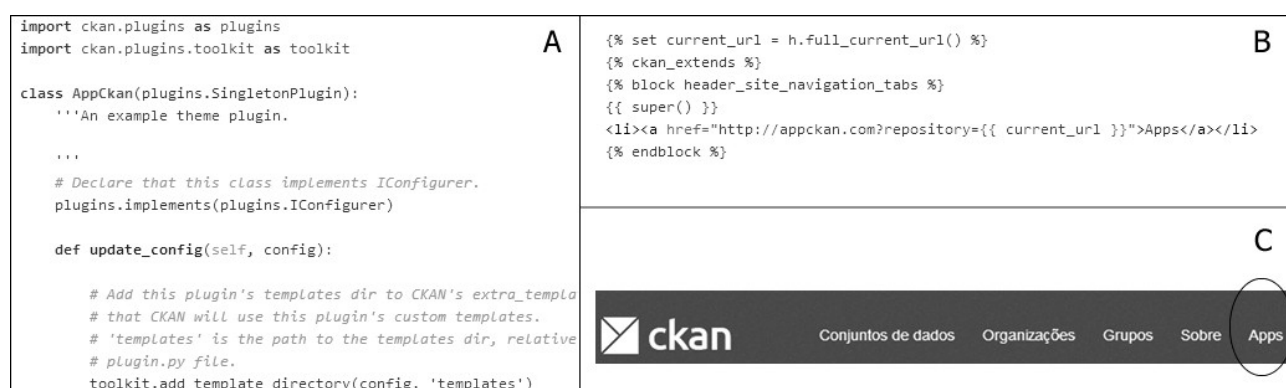


Figura 3: (A) Principal Classe da Extensão, (B) template que adiciona o botão e (C) o resultado final.

Aplicação APPCKan

Através do endereço <http://appckan.com> é possível acessar o APPCKan, uma aplicação que permite cadastrar as descrições dos aplicativos, de maneira a criar um repositório de aplicações que poderá ser facilmente consultado. Em seu desenvolvimento, foi usado o PHP (*PHP: Hypertext Preprocessor*) como linguagem de programação, a plataforma Wordpress como sistema de gerenciamento de conteúdo (CMS) e o MySQL como sistema de gerenciamento de banco de dados (SGBD).

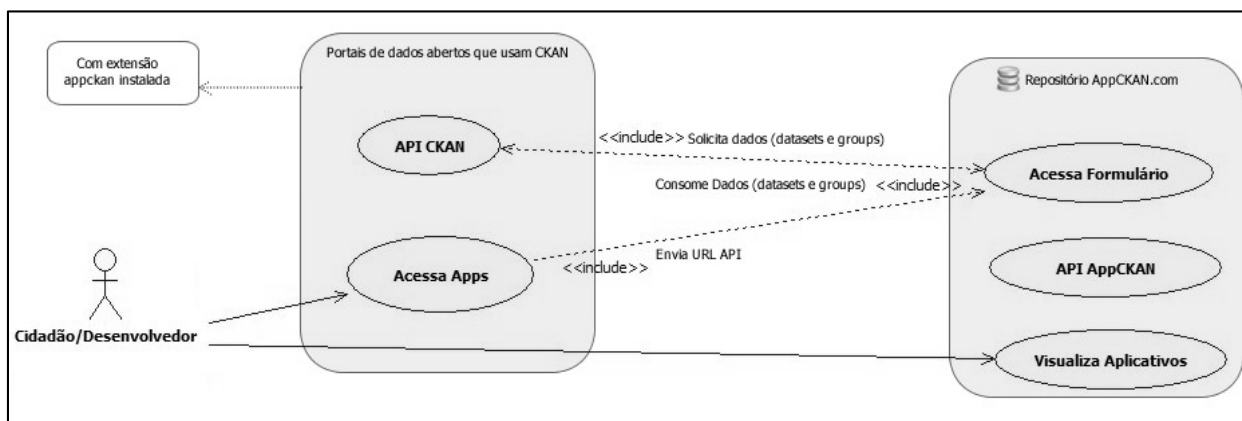


Figura 4: Visão geral da integração de um portal CKAN com o AppCKAN.com

Uma visão geral do repositório pode ser conferida na Figura 4. Quando o usuário acessar um portal de dados CKAN que tem a extensão *appckan* ativa, ele visualizará a opção “Apps” e ao entrar nela, será direcionado para o APPCKan.com, que receberá o endereço do portal de origem como parâmetro. Assim, o repositório carregará todos os *datasets* e os *groups* cadastrados no respectivo portal e exibirá o formulário para cadastro. Após o usuário preencher os dados do formulário de cadastro, o repositório processará e armazenará as informações no banco de dados. Disponibilizando, após, o cadastro para visualização. Sua interface e algumas de suas funcionalidades podem ser visualizadas na Figura 5.

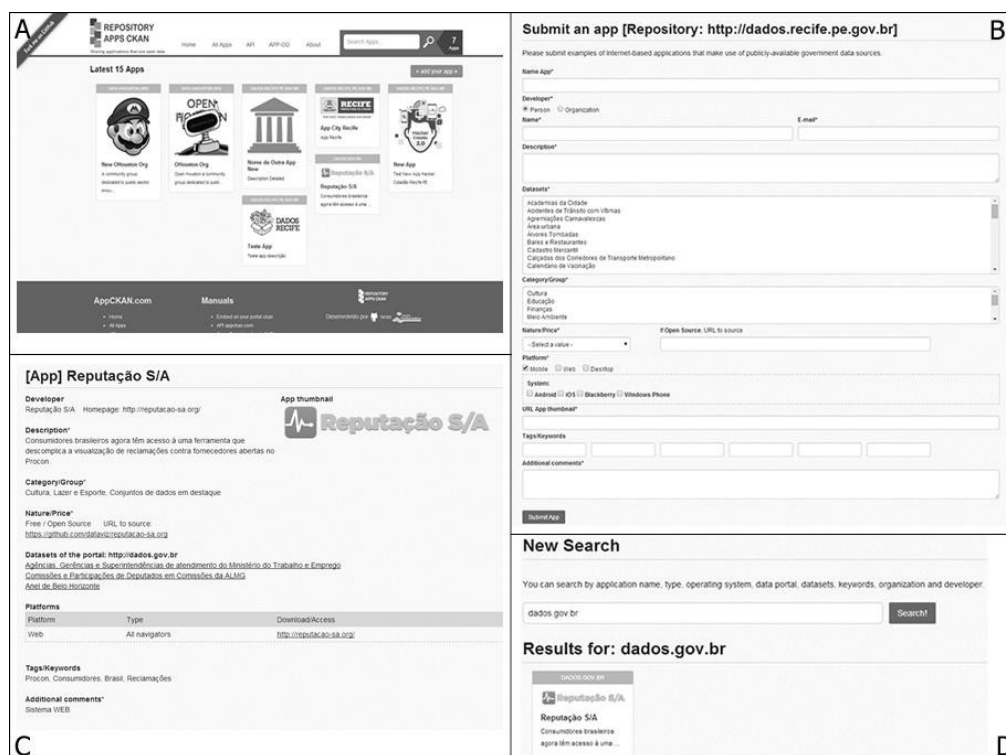


Figura 5: (A) Layout da aplicação com dados fictícios, (B) Formulário de cadastro, (C) Visualização de uma aplicação e (D) busca de aplicativos

O repositório ainda conta com uma API que permite listar todos os aplicativos e realizar filtros nos dados. Ela é disponibilizada no formato JSON e foi desenvolvida seguindo o mesmo padrão da API do CKAN, com três saídas: *help*, para descrever as possibilidades de acesso; *success*,

indicando se a consulta foi realizada com sucesso ou não; e o *result*, contendo um *array* com todos os resultados encontrados.

Por fim, esta primeira versão da aplicação APPCKan se preocupou em disponibilizar apenas as funcionalidades básicas para um ambiente deste tipo. Mas, com futuras integrações nos portais de dados abertos, ele poderá facilitar a busca de aplicativos de interesse dos usuários e na formação de comunidades, por exemplo. Somado a isso, ele também contribui fornecendo uma implementação do vocabulário APP-OD¹⁶. Seu código também está disponibilizado no Github a partir do endereço <https://github.com/lerao/portal-appckan>.

CONCLUSÕES

Este trabalho dividiu-se essencialmente em três partes: um estudo sobre a disponibilização de aplicativos que fazem uso de dados abertos; o desenvolvimento de um vocabulário para descrever tais aplicativos; e a implementação de um repositório online que permitisse o cadastro dessas descrições.

No estudo realizado, foram analisados 426 portais de dados abertos de 70 países, constatando que 28% listam ou tem uma seção para aplicativos e que apenas 17 portais, cerca de 4%, permitiam o cadastro de aplicativos. Além disso, que 78% dos portais pesquisados pertencem a 18 países, principalmente por terem aderido às iniciativas de dados abertos mais rapidamente. Contudo, o estudo mostrou que encontrar um aplicativo que resolva determinadas necessidades de um indivíduo não é uma tarefa trivial, tanto pelo fato de geralmente ser publicado em vários sites distintos, quanto por não serem encontrados nos principais portais de dados abertos. Também foi possível identificar que o CKAN por ter um bom nível de maturidade, ser *open source* e ter uma comunidade bastante ativa, está presente na maior parte dos portais.

A análise dos portais também permitiu elencar os principais dados solicitados nos formulários de aplicativos. Assim, eles foram sumarizados e usados como base para o desenvolvimento do vocabulário APP-OD, proposto para descrição de aplicações que fazem uso de dados abertos. Assim, esta contribuição visa agregar valor a um domínio que, até então, possui poucos vocabulários descritos na literatura.

Por fim, foi desenvolvido um repositório online para permitir o cadastro de descrições de aplicativos, o AppCKAN.com e uma extensão que permite fazer um link entre o portal CKAN e o repositório. O cadastro de aplicativos e a API disponibilizados foram desenvolvidos com base no vocabulário APP-OD. Espera-se, portanto, que ele venha a se tornar uma ferramenta útil para divulgação das aplicações que usam dados abertos, que possa ajudar na procura por aplicativos que satisfaçam determinadas necessidades das pessoas, ajude na formação de comunidades e discussões de novas ideias.

Este trabalho foi concebido como parte inicial de um estudo ainda maior na área. Dessa forma, algumas limitações deste trabalho são conhecidas e futuras ações são esperadas. Dentre

as elas, podemos destacar as seguintes: (1) Estruturar a captura de informações para gerar *feedbacks* quanto às operações que foram realizadas nos conjuntos de dados utilizados, como por exemplo, se foi necessário alterar os dados, se precisou unir mais de um *dataset* ou quais os dados que efetivamente foram usados; (2) Diminuir o número atual de requisições ao consumir os *datasets* e *groups*. Como estratégia de mitigação, espera-se que na próxima versão os dados consumidos na primeira requisição sejam armazenados localmente, para que uma segunda leitura seja mais rápida; (3) A extensão para o CKAN direciona apenas para o cadastro de aplicativos. Porém, espera-se que ela liste os aplicativos cadastrados do portal e que o cadastro seja de Forma transparente, sem necessidade de se deslocar para o repositório; (4) O repositório foi liberado considerando o status de desenvolvimento pré-alfa, sendo necessária a realização de testes e desenvolvimento de novas funcionalidades, como disponibilização de *reports* (gráficos e relatórios). Assim como, uma melhor visualização e apresentação geral do repositório. Além de ser necessário o desenvolvimento da interface em novos idiomas.

Em suma, este trabalho contribuiu para o desenvolvimento de uma abordagem para captura de informações sobre aplicações que fazem uso de dados abertos. Mesmo as iniciativas de dados abertos sendo recentes, elas têm contribuído bastante para aumentar a transparência, a colaboração e a participação dos cidadãos nas políticas e ações. Assim, podemos considerar as aplicações como o resultado dessa contribuição. Mas, além de serem desenvolvidas, também precisam ser bem divulgadas para serem usadas e não ficarem obsoletas. Portanto, com as futuras alterações no AppCKAN.com, também se espera que ele contribua de forma efetiva para essa divulgação, além de outros fatores mencionados ao longo deste trabalho.

REFERÊNCIAS

BERNERS-LEE, T.. **Linked Data**. 27 Jul 2006.

BRASIL. **Cartilha Técnica para Publicação de Dados Abertos no Brasil**. Brasília: MLTI; MP, 2012.

BRASIL. **Lei nº 12.527 de 18 de novembro de 2011**. Acesso à Informação Pública. Brasília, 18 Nov 2011.

CKAN. **Writing extensions tutorial**. 2013.

CÓRDOVA, Y.. **Open Data for Development. Gerenciamento do ciclo de abertura de dados**. 2013.

CUNHA, D. R. B.; SOUZA, D. Y.; LÓSCIO, B. F.. **Linked data: da web de documentos para a web de dados**. Teresina: SBC, 2011.

DUTRA, C. C.; LOPES, K. M. G.. Dados abertos: uma forma inovadora de transparência. In: Congresso CONSAD de Gestão Pública, Painel 20/075, 6. **Anais**. Brasília, 2013.

EAVES, D.. **The Three Laws of Open Government Data**. Vancouver, 2009.

FONSECA FILHO, B. E.; SOBOTA, L. O.. Política Externa para a Sociedade da Informação: a contribuição da diplomacia para a democracia na era digital. In: CONGRESSO INTERNACIONAL SOFTWARE: DADOS ABERTOS PARA A DEMOCRACIA NA ERA DIGITAL, 6. **Anais**. Brasília, 2011.

GHOSH, R. A.. **Economic impact of open source software on innovation and the competitiveness of the information and communication technologies (ICT) sector in the EU**. Maastricht: UNU-MERIT, 2006.

JOHNSON, C.. **Build Communities Not Contests**. The Information Diet. 2010.

JUNAR. **About: Junar open data platform**. 2014.

LOUTAS, N.; VARITIMOU, A.; PERISTERAS, V.. Unraveling the mystery of Open Government Data Apps. In: WORKSHOP ON USING OPEN DATA: POLICY MODELING, CITIZEN EMPOWERMENT, DATA JOURNALISM, 1. **Anais**. Brussels: W3C, 2012.

MIRANDA, C.. **Abra com simplicidade**: desmistificando os dados abertos. Open Data for Development, 2013.

MOREIRA, G. A.; LÓSCIO, B. F.. **GovPEX3**: uma ferramenta para publicação das despesas do Governo de Pernambuco em RDF. Recife: UFPE, 2013.

OKF. **What is Open?**, 2007.

OLIVEIRA, E.; PINTO, C. E. C.; AZEVEDO, M. M.. Dados abertos governamentais: uma forma da sociedade contribuir e participar na discussão e resolução de problemas. In: WORKSHOP DE PÓS-GRADUAÇÃO E PESQUISA, 8. **Anais**. 2013.

TCE-RS. Por que dados abertos? Rio Grande do Sul: 2013.

W3C. **Manual dos dados abertos**: desenvolvedor. 2011.

W3C. **Data Catalog Vocabulary**: DCAT. 2012.

W3C. **Data usage notes**: DWBP. 2013.

W3C. **What is Inference?**. 2014.