

PROFISSIONAL DA INFORMAÇÃO NO CONTEXTO DOS DADOS ABERTOS: O USO DO CKAN PARA A DISPONIBILIZAÇÃO E A ORGANIZAÇÃO DE DADOS

PROFESIONAL DE LA INFORMACIÓN EN EL CONTEXTO DE DATOS ABIERTOS: EL USO DE CKAN PARA LA DISPONIBILIDAD Y ORGANIZACIÓN DE DATOS

Caio Saraiva Coneglian*
José Eduardo Santarem Segundo**

RESUMO:

Introdução: O desenvolvimento das tecnologias está promovendo um aumento na geração de dados. Neste sentido, há diversas áreas que necessitam promover a abertura de suas informações, em destaque empresarial, institucional, governamental e acadêmica. O profissional da informação pode ter uma atuação neste cenário, porém é necessário que sejam utilizadas tecnologias que contribuam para isso, como a ferramenta CKAN, que promove a disponibilização de dados abertos.

Objetivo: Discutir o uso do CKAN por arquivistas e por bibliotecários para a organização e a representação de dados em formatos abertos.

Metodologia: Qualitativa de caráter exploratório, com embasamento teórico acerca de dados abertos, do sistema CKAN e dos profissionais da informação, e teórico aplicado, pois discutimos o uso do sistema CKAN para as funções dos profissionais da informação, realizando testes em um sistema implantado para este trabalho.

Resultados: Por meio de discussões teóricas e testes dentro da ferramenta, demonstramos a relação do CKAN com as atividades dos profissionais da informação, e como tal ferramenta pode ser utilizado para a execução destas atividades.

Conclusões: O CKAN se mostra como um sistema adequado para a organização e representação de dados abertos tanto para arquivistas quanto para bibliotecários, fornecendo funcionalidades que facilitam a publicação de dados abertos.

Palavras-chave: Dados abertos. Profissional da informação. Arquivista. Bibliotecário. CKAN.

*Mestrando pelo Programa de Pós-Graduação em Ciência da Informação da UNESP Marília. Bacharel em Ciência da Computação pela UNIVEM. E-mail: caio.coneglian@gmail.com

**Docente do Departamento de Educação, Informação e Comunicação da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da USP. Docente do Programa de Pós-Graduação em Ciência da Informação da UNESP Marília. Doutor pelo Programa de Pós-Graduação em Ciência da Informação da UNESP Marília. E-mail: santarem@usp.br

1 INTRODUÇÃO

Os estudos com dados se mostram cada vez mais pertinentes por vivermos em um momento em que a geração de informação cresce de forma exponencial. Os dados se tornaram um elemento essencial para todos os tipos de organizações, além de ser fundamental dentro da Ciência. Neste cenário, discussões acerca dos dados abertos começaram a ser feitas, principalmente no que diz respeito às políticas e aos ferramentais para a abertura das informações.

Assim, a organização chamada de *Open Knowledge Foundation* passou a ter um papel relevante nestas discussões que promove o conhecimento aberto em diversas esferas. Tal organização apresenta diversas normativas, além de estar à frente do desenvolvimento e da manutenção do *Comprehensive Knowledge Archive Network* (CKAN), que tem como função ser uma ferramenta para a disponibilização de dados abertos. (OPEN KNOWLEDGE BRASIL, 2016).

Os dados abertos estão tendo alta visibilidade em pelo menos dois cenários, referente às organizações, às empresas e aos governos que necessitam estar adequados à legislação e por pressão popular precisam abrir seus dados a sociedade e os dados de pesquisa, que apresenta uma tendência de abertura, visando o aprimoramento da Ciência como um todo. Uma consequência destes processos é a existência de uma quantidade cada vez maior de instituições disponibilizando seus dados, tendo a necessidade de ter meios para estruturar tais informações para melhorar a recuperação posterior.

Neste sentido, os profissionais da informação podem trazer contribuições principalmente da organização e da representação destas informações. Especialmente os arquivistas e os bibliotecários podem ter um papel central no que cerne a implantação de ambientes informacionais para a disponibilização dos dados abertos, sendo o CKAN uma ferramenta que colabora com isso. Vale destacar que o CKAN é utilizado como o principal sistema para a disponibilização de conjuntos de dados (*datasets*) do *Linked Open Data* (LOD), que visa disponibilizar abertamente dados em formatos semânticos com ligações entre eles, promovendo a descoberta de informações pelos usuários.

Questiona-se assim, como o CKAN pode auxiliar as funções que compete aos profissionais da informação na organização e na representação de dados abertos,

bem como as funcionalidades que tal sistema fornece para o trabalho deste profissional.

Desta forma, esta pesquisa objetiva discutir o uso do CKAN por arquivistas e por bibliotecários para a organização e a representação de dados em formatos abertos. Para tal, utilizou-se uma metodologia qualitativa de caráter exploratório, uma vez que se buscou na literatura embasamento acerca de dados abertos, do sistema CKAN e dos arquivistas e dos bibliotecários, e teórico aplicado, pois discutimos o uso do sistema CKAN para as funções dos profissionais da informação, realizando testes em um sistema implantado para este trabalho.

2 DADOS ABERTOS E LIGADOS

A evolução das Tecnologias da Informação e Comunicação fez com que nas mais distintas tarefas realizadas por uma pessoa, fossem utilizados equipamentos eletrônicos. Essa mudança gerou um aumento substancial na geração de dados, que provocou mudanças em todas as áreas do conhecimento e instituições, como empresas, governo e academia.

Neste contexto, movimentos como *Big Data*, que se refere a grandes quantidades de dados gerados rapidamente e nos mais diversos formatos, e *Big Science*, áreas como astronomia, biologia e física que criam grandes estruturas para o tratamento e processamento de dados, começaram a olhar para os dados como um instrumento essencial nas análises e nas pesquisas, utilizando-os, entre outras coisas, para a determinação de comportamentos dos consumidores e para descobertas científicas. A partir disso, disciplinas nas áreas do conhecimento da Ciência da Computação e da Ciência da Informação como a Ciência de Dados, a Curadoria Digital e o *Big Data Analytics* foram criadas, passando a estudar os fenômenos relativos aos dados.

Uma parte fundamental destes estudos diz respeito a questão do acesso à informação, pois tão importante quanto gerar, armazenar e processar é ter mecanismos que tornem o acesso aos dados eficiente. Santarém Segundo (2015, p. 221) relata que “As instituições, sejam elas públicas ou privadas, tem investido na organização e no acesso a informação como o grande diferencial na tomada de decisão em várias de suas instâncias.”

A afirmação do autor indica a importância que as informações e os dados possuem para as instituições, sejam elas governamentais ou não. No âmbito das organizações governamentais, há uma tendência crescente relativa a disponibilização dos dados abertamente, pautadas por leis e por pressões populares, na busca de deixar mais transparente as ações governamentais.

A organização *Open Knowledge Foundation* apresenta um papel importante neste cenário de promover a abertura dos dados. A *Open Knowledge Foundation* foi concebida em 2004 em Cambridge sendo uma organização sem fins lucrativos. Na página Web do capítulo brasileiro desta organização, ela é definida como:

A Open Knowledge Brasil é uma organização nacional apartidária, sem fins lucrativos, que utiliza e desenvolve ferramentas cívicas, faz análises de políticas públicas, trabalha com jornalismo de dados e promove o conhecimento livre para tornar a relação entre governo e sociedade mais transparente e para que haja uma participação política mais efetiva e aberta. (OPEN KNOWLEDGE BRASIL, 2016).

Buscando nortear os projetos que buscam disponibilizar seus dados abertamente, a *Open Knowledge Foundation* criou o Manual dos Dados Abertos (*The Open Data Handbook*). Neste documento, dados abertos são definidos como “[...] dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras.” (DIETRICH et al., 2009).

Por meio desta definição, verifica-se que o conceito de dados abertos abrange diversos aspectos que devem ser respeitados, que vão desde questões relativas a políticas de acesso até questões como regras para atribuir fontes e compartilhamento. Vale destacar ainda a questão relativa ao acesso por qualquer pessoa, indicando a forte relação que existe entre os dados abertos e a transparência de acesso.

Assim, Santarem Segundo (2015) complementa essa questão, inserindo a necessidade da sociedade consumir os dados. O autor afirma que ao se referir a sociedade devemos pensar nas diversas esferas organizacionais, como academia, jornalismo, iniciativa privada, organizações não-governamentais, além de pessoas isoladamente, que possuem o interesse de acessar os dados independente do objetivo desta ação.

Uma questão levantada por Santarem Segundo (2013) demonstra um dos principais erros ao se publicar dados, que é a utilização de formatos proprietários na disponibilização destes. Neste sentido, o autor afirma que “A utilização de formatos em que seja necessária a utilização de ferramentas do pacote Office da Microsoft, por exemplo, pode inviabilizar o acesso aos dados pela comunidade.” (SANTAREM SEGUNDO, 2013, p. 35-36). O autor indica uma outra questão que prejudica consideravelmente o acesso aos dados, que é a disponibilização em formato *Portable Document Format* (PDF), pois inviabiliza a leitura destes por mecanismos automatizados.

As duas características comprometem o acesso aberto, uma vez que caso seja utilizado formatos proprietários, há uma dependência do uso de ferramentas construídas por determinadas empresas que podem limitar o acesso, enquanto o uso do PDF dificulta, e em alguns casos impede, o acesso aos dados por agente computacionais, pelas dificuldades existentes na implementação de mecanismos que interpretem textos em PDF.

Estas questões convergem com os pontos mais importantes elencados pelo Manual de Dados Abertos, citado anteriormente, em que os autores resumem em três principais pontos: (i) a disponibilização e acesso, demonstrando a necessidade da facilidade no acesso aos dados, permitindo preferencialmente com que os dados possam ser baixados pela internet; (ii) reutilização e redistribuição, indicando que os dados devem ser fornecidos para que possam ser reutilizados e redistribuídos e; (iii) participação universal, todas as pessoas devem ser capazes de acessarem, reutilizarem e redistribuírem os dados, de forma que não pode haver nenhum tipo de discriminação em sua disponibilização. (DIETRICH et al., 2009).

Uma das principais atuações dos estudos de Dados Aberto diz respeito aos dados governamentais, que envolve questões como a transparência e a granularidade. Neste sentido, há diversos questionamentos que abordam o papel da sociedade no acesso aos dados, além de como as tecnologias podem aprimorar o controle da sociedade sobre o estado. Rodrigues e Sant’ana reafirmam tal questão, ao indicar que: “A sociedade pode e deve exercer um controle sobre o poder do Estado e; o Estado retroalimenta-se, ganhando novos mecanismos para perceber melhor as necessidades dos cidadãos.” (RODRIGUES; SANT’ANA, 2012, p. 4).

Por meio das afirmações feitas pelos autores, verifica-se que as tecnologias podem ser um meio de promover esse controle. Lévy (2015) em seu livro “A Inteligência Coletiva: por uma antropologia do ciberespaço” discorre em alguns momentos sobre o poder que o ciberespaço e as tecnologias possuem no que concerne participação popular nas decisões do estado. O autor ainda relata que: “O uso socialmente mais rico da informática comunicacional consiste, sem dúvida, em fornecer aos grupos humanos os meios de reunir suas forças mentais para constituir coletivos inteligentes e dar vida a uma democracia em tempo real.” (LÉVY, 2015, p. 64).

No âmbito governamental, os dados abertos apresentam um papel importante para promover as forças mentais destacadas por Lévy, além de indicar um caminho que grande parte dos países estão adotando, com destaque na Lei de Acesso à Informação (nº 12.527/2011) no que se refere ao Brasil. Santarém Segundo (2013, p. 1) destaca que esta lei “[...] tem despertado um conjunto de ações e, principalmente, de pesquisas com a intenção de abrir caminhos para que o povo brasileiro tenha acesso às informações governamentais de forma mais ampla e clara.”

A partir destas reflexões, verifica-se que os dados abertos governamentais em consonância com as tecnologias estão a apresentar um papel bastante relevante na sociedade, tendo como uma das principais áreas de estudos a Ciência da Informação, com destaque no estudo de tecnologias que podem aprimorar a participação dos profissionais da informação neste contexto. Há também um segundo campo de estudo dos dados abertos, que trata dos dados de pesquisa.

O conceito de dados abertos de pesquisa apresenta uma forte relação com questões relativas a Ciência Aberta, que busca deixar a ciência mais transparente, mais colaborativa e mais eficiente (SAYÃO; SALES, 2014). Neste sentido, os dados abertos apresentam um papel importante para a difusão da ciência, além de tornar a ciência aberta. Sayão e Sales (2014, p. 82) ao discorrer sobre dados abertos de pesquisa, relatam que

[...] pode-se pressupor que o reuso e o compartilhamento de dados e informações, num ambiente de pesquisa caracterizado pela pluralidade de visão sobre esses recursos, abrem a possibilidade de se conceituar formas inéditas de agregações abstratas de produtos de pesquisa que sejam portadores de interpretações específicas,

criando, dessa forma, novos constructos intelectuais que possuam os atributos mínimos dos recursos informacionais, ou seja, possam ser identificados e tenham sua autoria reconhecida. (SAYÃO; SALES, 2014, p. 82).

A partir das afirmações realizadas, verifica-se que os dados abertos trazem uma grande contribuição para a Ciência, possibilitando indiretamente a geração de novos conhecimentos por meio do compartilhamento. Cada vez mais os dados abertos de pesquisa estão ganhando destaque e se tornam uma realidade dentro do meio acadêmico, devendo haver uma preocupação em questões relativas ao armazenamento, à organização, à representação e à recuperação destes dados.

Uma outra questão de bastante destaque no que diz respeito aos dados abertos, é *Linked Open Data*. O *Linked Open Data* nasceu a partir das recomendações propostas por Tim Berners-Lee em 2006, acerca do *Linked Data*, que previa uma forma de publicação dos dados, utilizando as tecnologias e os conceitos da Web Semântica. (BERNERS-LEE, 2006).

Atualmente, a iniciativa do *Linked Open Data* contém um grande número de *datasets* interligados, que promove a descoberta de informações pelos usuários, tendo seus dados representados em formatos processáveis por máquinas, como *Resource Description Framework* (RDF) e *eXtensible Markup Language* (XML). Assim, a Web Semântica que foi proposta em 2001 e tinha como uma de suas bases com que os dados estivessem estruturados e contextualizados para que agentes computacionais fossem capazes de compreender as informações, tem o *Linked Open Data* como uma de suas principais iniciativas (BERNERS-LEE; HENDLER; LASSILA, 2001).

O *Linked Open Data* se mostra como uma iniciativa que vai ao encontro das recomendações de Dados Abertos, contribuindo na disponibilização das informações de forma que estes possam ser recuperados e reutilizados futuramente. Para que as práticas de *Linked Data* sejam aplicadas é necessário que existam ferramentais e portais que permitam a disponibilização de bases de dados na Web, as deixando abertas e acessíveis para usuários humanos e não-humanos.

Dentro dessa perspectiva e com a intenção de criar uma plataforma que possibilite a disponibilização de dados acessíveis, a *Open Knowledge Foundation*, iniciou o desenvolvimento do CKAN em 2006. Na sequência será tratado acerca do

CKAN que apresenta um papel fundamental para o *Linked Data* e para os dados abertos.

3 CKAN

O CKAN é uma ferramenta para a construção de portais de dados abertos, que auxilia na publicação e no gerenciamento de coleções de dados, sendo uma plataforma *open source*, com uma comunidade de desenvolvedores que busca corrigir falhas e inserir novas funcionalidades periodicamente. Esta ferramenta tem como funções primárias os processos de publicação, de compartilhamento, de descoberta e de uso dos dados. Os principais utilizadores desse sistema são organizações, companhias e governos que desejam disponibilizar suas informações abertamente (CKAN, 2015a).

O desenvolvimento do CKAN foi idealizado pela *Open Knowledge Foundation* devido a uma das diretrizes dessa fundação tratar do desenvolvimento, do suporte e da promoção de ferramentas que busquem facilitar a criação, o acesso e a disseminação de conhecimentos (WINN, 2013).

O produto principal tratado dentro dessa plataforma são dados, assim, é necessário que exista uma organização do como estes devem estar estruturados. O CKAN organiza todos os dados em *datasets* (conjuntos de dados), sendo que cada *dataset* contém informações acerca de um determinado contexto, como por exemplo informações governamentais de gastos com obras públicas. A partir dessa organização, os usuários que fizerem buscas, recebem as informações individualmente, e não todo o conjunto. Basicamente, um *dataset* contém dois tipos de informações: os metadados que descrevem informações dos dados, como data de criação, criador, título e; conjuntos de recursos, que são os dados em si, podendo ser os recursos de qualquer formato, como PDF, *Comma-Separated Values* (CSV), Excel, RDF, entre outros, ou um recurso pode ser também um *link* que leva a algo na Web (CKAN, 2013).

As principais funcionalidades que a plataforma CKAN apresenta são (CKAN, 2015b):

- A publicação e gerenciamento dos dados, em que estes podem ser inseridos por meio de uma interface Web, pela utilização de uma API ou por planilhas customizadas;

- O processo de *harvesting*, em que podem ser coletados dados de diversas outras fontes que as organizações mantêm ativas;
- Busca e descoberta de informações, em que a plataforma mantém um sistema de buscas baseados nos buscadores tradicionais, e que a utilização de metadados na descrição dos dados melhora substancialmente na recuperação dos registros;
- Recursos geoespaciais para aperfeiçoar a busca e a descoberta dos dados;
- Ferramentas de comunidades, em que é possível que os usuários se interajam dentro da plataforma para haver comunicação e colaboração nos dados;
- A visualização, que permite diversas funcionalidades com o intuito de verificar e entender as características dos dados, e permitir com que o usuário consiga explorar mais profundamente os *datasets*;
- Customização, permitindo que sejam inseridos diversos temas para se adequar as necessidades de cada organização;
- Armazenamento, em que o CKAN permite opções de armazenamento seguro dos dados;
- Histórico, onde todas as ações que são realizadas com dados são armazenadas em um histórico, que pode ser consultado posteriormente;
- Extensões, que incluem diversas outras funcionalidades ao sistema;
- Federação, devido ao processo de *harvesting* é possível que seja criado uma federação com diversos nós do CKAN, que compartilham dados entre eles e;
- API, em que é disponibilizado um servidor de requisições para buscas e acesso às informações dos *datasets* contidos no servidor CKAN.

Santarem Segundo e Faria (2013) destacam os diversos benefícios que o CKAN oferece, citando o fato de que a ferramenta ser *open source* e as possibilidades de ser executado *harvesting*, auxilia substancialmente a implantação e uso desta. Os autores dizem ainda, que tal plataforma expande as questões de ser um simples depósito de dados, pela utilização de diversos conceitos de publicação da informação em ambientes abertos, sendo que o CKAN é atualmente a opção mais adequada para a publicação de dados em formato aberto disponível para uso.

Um outro destaque do CKAN é o padrão de metadados utilizado para a representação dos objetos digitais. Tal padrão é composto pelos seguintes elementos: título; identificador único; descrição; histórico de revisão; visualização de dados; campos extras; licença; *tags*; grupos e; múltiplos formatos (CKAN, 2015b). Santarem Segundo e Faria (2013) destacam que o padrão de metadados do CKAN apresenta uma preocupação em manter a compatibilidade com outros formatos, que pode ser visto por exemplo pela “[...] relação estreita com o padrão Dublin Core.” (SANTAREM SEGUNDO; FARIA, p. 11, 2013). Os autores destacam ainda que este padrão do CKAN foi desenvolvido para a própria ferramenta, porém podendo ser facilmente adaptado.

As estruturas de representação e de organização da informação utilizadas pelo CKAN serão melhores exploradas a seguir, em que são apresentados os resultados e as discussões desta pesquisa.

4 RESULTADOS E DISCUSSÕES

A forma de disponibilização de dados abertos se mostra como uma discussão recorrente, no que diz respeito a quais procedimentos devem ser adotados, bem como a identificação de plataformas que possibilitam a disponibilização, a organização e a representação adequada dos dados. Neste sentido, os profissionais da informação mostram-se como um elemento essencial no contexto da publicação de dados abertos e de sua manutenção.

A existência de uma plataforma que seja capaz de atender os requisitos para a organização e a representação dos dados, juntamente com os profissionais capazes de executar estas tarefas, torna o processo de recuperação dos dados mais eficiente, ao passo que estes estarão passíveis de serem encontrados com facilidade. A partir desta ótica, o trabalho do profissional da informação no cenário dos dados abertos é essencial, pois é por meio de suas habilidades, especialmente no que tange a organização e a representação das informações, que os dados poderão ser recuperados, possibilitando com que usuários sejam capazes de acessar e de (re) utilizar tais dados.

Nesta seara, o CKAN surge como um ferramental de extrema importância, pois como relatado, esta plataforma foi proposta pelo *Open Knowledge Foundation*, estando de acordo com as principais diretrizes dos dados abertos. O CKAN

composto por diversas funcionalidades consegue atender as necessidades dos profissionais da informação em sua tarefa referente aos dados abertos. Destaca-se entre outras funcionalidades, a criação de diversas organizações, a descrição dos dados publicados e as formas de recuperação da informação.

Identifica-se dentro dos profissionais da informação, dois que apresentam uma atuação mais ampla e profunda no contexto dos dados abertos, o arquivista e o bibliotecário. Para demonstrar como estes profissionais podem atuar utilizando a plataforma CKAN, realizou-se a instalação desta. Para tal, seguiu-se o manual de instalação do próprio sistema¹. Posteriormente foram realizados testes buscando demonstrar como cada um dos profissionais poderiam atuar utilizando o CKAN. Para tal, na sequência foram criadas duas subseções para realizar a discussão de como os profissionais da arquivologia, primeiramente, e da biblioteconomia na sequência, poderiam utilizar esta tecnologia.

4.1 Arquivista

O profissional arquivista tem em suas atribuições funções relativas a organização das informações nos mais variados contextos, necessitando utilizar sistemas de informação para aprimorar seu trabalho. Neste sentido, Ribeiro (2005, p. 57) relata que:

[...] o arquivista da era pós-custodial, seja ainda designado desta forma ou venha a ter um título mais pós-moderno, será fundamentalmente um Profissional da Informação, com uma formação de base, suficientemente sólida e abrangente, que lhe permitirá exercer funções em qualquer contexto orgânico produtor/manipulador de fluxo informacional. Mas, estará devidamente preparado para actuar, também, em contextos de algumas especificidades, seja no âmbito de sistemas de informação organizacionais ou em serviços de arquivo especializados, porque o seu referencial teórico reporta sempre ao campo do saber que lhe dá identidade - a Ciência da Informação (RIBEIRO, 2005, p. 57).

Por meio da definição apresentada, verifica-se que o arquivista está se adaptando ao atual momento da sociedade, em que existe um aumento exponencial na quantidade de informações, havendo a necessidade constante de utilizar as novas tecnologias. O arquivista enquanto profissional da informação tem como uma

¹ Instalação do CKAN. Disponível em:
<<http://docs.ckan.org/en/latest/maintaining/installing/index.html>>. Acesso em: 18 jun. 2016.

de suas funções lidar com o fluxo informacional, atuando nos mais variados tipos de arquivos.

Neste sentido, a busca por profissionais que trabalhem para a disponibilização de dados abertos nas instituições tem aumentado, seja empresas, organizações governamentais e não-governamentais, municípios, estados e federação, entre outras. O arquivista se mostra preparado para atuar nestas áreas, por ter em sua formação os conceitos para a organização de um arquivo, estruturando-o para permitir uma posterior recuperação dos dados conforme a necessidade dos usuários.

A estruturação dos dados e *datasets* é um processo fundamental para a implementação, a disponibilização e a divulgação dos dados abertos. Assim, o uso do CKAN pelo profissional arquivista se mostra como uma solução para a publicação de dados abertos das entidades citadas.

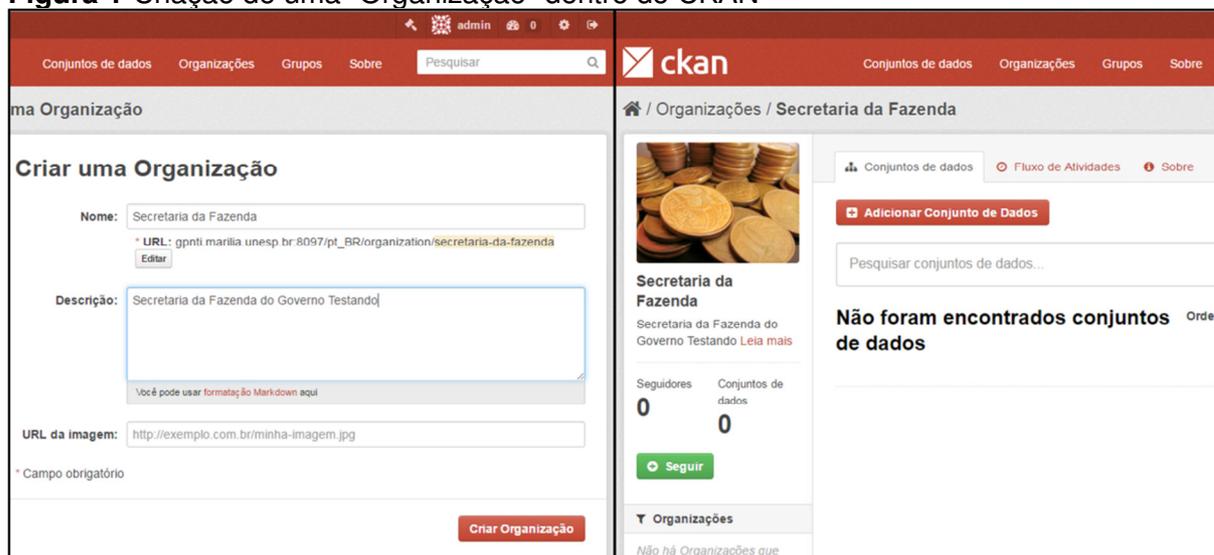
Neste âmbito, uma primeira funcionalidade que fornece meios para realizar a organização dos dados é a possibilidade de criar “Organizações” que estabelece separações entre diversos grupos em uma instituição. Um exemplo de criação destas organizações, seria a separação dentro de municípios, de estados e da federação, em secretarias, em ministérios e em institutos, e se tratando de empresas e de organizações (governamentais ou não) esta divisão poderia ser feita em departamentos, em setores e em diretorias.

A estrutura de divisão nas chamadas organizações promove a transparência dos dados, além de favorecer a recuperação posterior, uma vez que haverá divisões claras sobre a origem dos mesmos. O processo de criação de organizações se mostra bastante facilitado, necessitando a criação por um usuário administrador, que deverá inserir somente o nome daquela organização (secretaria, departamento, ministério...), uma URL que apontará para o espaço criado e uma breve descrição.

Buscando demonstrar o processo de criação de uma organização, realizou-se esse processo no CKAN. A Figura 1 ilustra os ambientes que compõem o processo de criação de uma organização. Na esquerda da figura está apresentado um formulário para a criação da organização, que como relatado exige o nome, a URL, a descrição e uma figura que pode ser inserida para identificar aquela determinada organização. A direita da figura, contém o ambiente inicial da organização criada,

que apresenta as informações descritas no processo de criação da organização, além de possibilitar a inserção e a visualização de conjuntos de dados.

Figura 1-Criação de uma “Organização” dentro do CKAN



Fonte: Elaborado pelos autores.

A partir da criação de uma ou mais organizações do CKAN, é possível que sejam feitas as inserções dos conjuntos de dados. Tal processo apresenta o maior nível de complexidade, pois deverá ser feito de forma que promova uma representação expressiva das informações. Vale destacar ainda que o processo de disponibilização deverá atender os princípios de transparência, de acordo com a legislação vigente. Neste sentido, torna-se necessário que existam meios de sinalizar os formatos dos dados, para que caso hajanormativas no que tange a disponibilização em formatos processáveis por máquinas, uma instituição possa apontar claramente se está atendendo a elas.

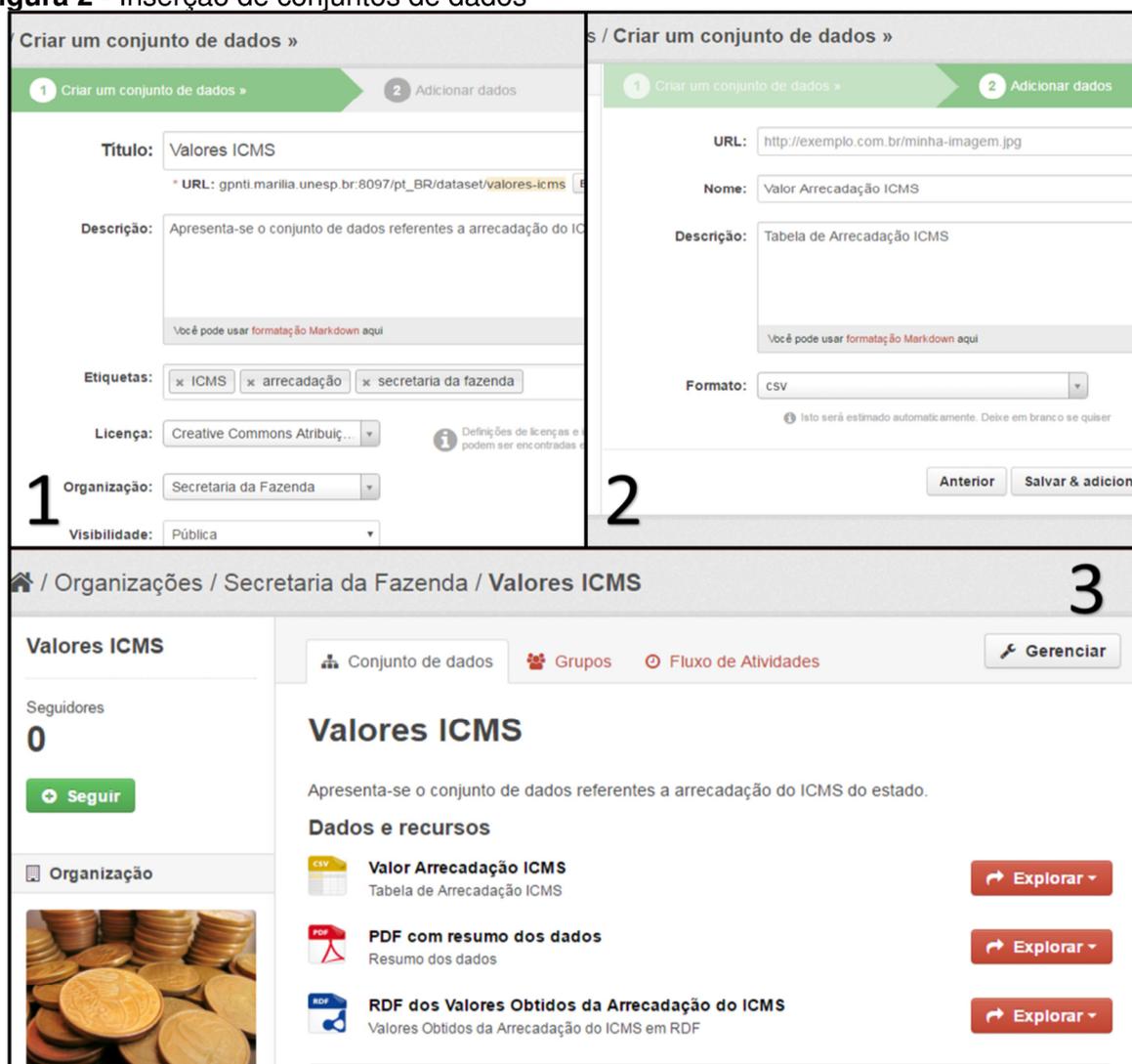
O processo de inserção de *datasets* ainda exige uma série de informações que promove uma identificação clara das características dos dados, dos autores e de metadados para realizar uma descrição adequada das informações. Desta forma, um conjunto de dados deverá estar necessariamente vinculado a uma organização, devendo estar representado: o título, uma descrição, *tags* (etiquetas), a licença (por exemplo *Creative Commons Atribuição*), a visibilidade (pública ou privada), a fonte, informações do autor e do mantenedor, além de campos customizados, que podem ser utilizados, entre outras coisas, para a inserção de outros metadados.

Posteriormente a inclusão destas informações, são inseridos os dados propriamente dito. Neste passo, deve-se apontar uma URL que contém o endereço

em que eles estão armazenados, um nome para esses dados, uma descrição e o formato em que os mesmos estão disponibilizados (por exemplo, PDF, CSV e RDF). Cabe duas ressalvas desse momento, primeiramente que os dados não ficarão armazenados propriamente na ferramenta CKAN, mas sim em outros servidores, ficando a cargo do CKAN o gerenciamento dos conjuntos de dados, e a necessidade de apontar os formatos deles, pois os formatos serão expostos aos usuários no momento da recuperação da informação.

O processo de inserção dos dados pode ser repetido inúmeras vezes, representando as diversas informações que podem compor um determinado conjunto. Neste passo do processo, o arquivista necessita estar atento a diversas questões que envolvem a disponibilização de dados como, a licença e o formato dos mesmos.

Figura 2 - Inserção de conjuntos de dados



Fonte: Elaborado pelos autores.

Desta forma, realizou-se uma demonstração das principais tarefas que perpassam na função da inserção dos *datasets* no CKAN, buscando demonstrar como o arquivista poderá inserir tais informações no sistema. A Figura 2 engloba os passos que poderão ser feitos dentro do sistema.

A Figura 2 apresenta três separações que indica distintos ambientes de interação que o usuário terá contato ao inserir conjuntos de dados. No primeiro ambiente o usuário insere às informações que descrevem o *dataset*, já no segundo ambiente, serão inseridos os dados em si, destacando que o usuário poderá repetir este processo inúmeras vezes, e por fim, a tela de número três indica a apresentação do *dataset* após a sua inserção. Esta tela de apresentação insere ícones que indicam os tipos de dados inseridos, facilitando a localização por parte dos usuários, além disso há a opção de explorar os dados, que fornecerá ao usuário um *link* que conduz ao local em que eles estão armazenados, além de em alguns casos permitir a sua visualização, como o CSV que monta uma tabela e o PDF que abre um leitor de PDF do próprio navegador.

Como destacado anteriormente, o trabalho com dados abre uma gama de possibilidades para os arquivistas, sendo necessário que o profissional desta área encontre meios que facilite a execução destas tarefas. Pelas demonstrações e explanações realizadas, verifica-se que o CKAN fornece um caminho para o auxílio nestas funções, especialmente as que dizem respeito a dados abertos. As funcionalidades do CKAN permitem uma descrição clara, podendo ser usado por distintas instituições, de modo a atender as legislações e a transparência pública, podendo ser o arquivista o profissional a desempenhar esta tarefa.

4.2 Bibliotecário

O bibliotecário tem dentre as suas principais especialidades a representação e a organização da informação, permitindo com que as informações possam ser recuperadas futuramente. Neste sentido, diversas pesquisas apontam da importância do profissional bibliotecário no que concerne aos novos ambientes informacionais digitais, como a Web e a Internet, exigindo uma profunda conexão entre o bibliotecário e a informática na busca de encontrar soluções para uma

melhor organização dos dados contidos na Web. Assim, Barbosa já no ano de 1998 aponta a importância da biblioteconomia nos novos ambientes digitais ao relatar que:

A aplicação do conhecimento tradicional da biblioteconomia em novos ambientes e contextos não apenas irá ajudar a desenvolver novas oportunidades de mercado para os profissionais, mas também irá contribuir decisivamente para o efetivo gerenciamento do acervo de conhecimentos da sociedade (BARBOSA, 1998).

A evolução dos ambientes levou a questionamento acerca dos processos de organização e de representação dos dados, em destaque aos dados de pesquisa que estão tendo um papel central dentro da academia. Desta forma, cada vez mais a preservação dos dados de pesquisa se torna tão importante quanto o armazenamento do texto escrito de uma tese, de uma dissertação ou de um artigo científico. A partir disto, o bibliotecário respaldado por suas competências surge como um agente central neste processo. Costa e Cunha validam esta questão ao dizer que o profissional da informação, principalmente o bibliotecário no cenário dos dados abertos de pesquisa precisam “[...] compreender esse novo cenário de dilúvio de dados e como podem contribuir para uma gestão eficiente da informação.” (COSTA; CUNHA, 2014, p. 201).

Uma outra questão a ser levantada, diz respeito às bibliotecas serem um dos ambientes que armazena entre outros objetos, as teses, as dissertações e os artigos científicos, que como consequência dessa evolução na ciência em que os dados necessitam também estar disponibilizados, fica a cargo da biblioteca e do bibliotecário encontrar formas que permitam o armazenamento e a recuperação posterior destes dados. Neste sentido, torna-se necessário que os bibliotecários identifiquem meios que possibilite a vinculação entre os objetos digitais textuais com os dados oriundos das pesquisas.

A partir destas constatações, verifica-se a importância de ferramentas tecnológicas que contribuam para que os bibliotecários possam trazer a sua expertise no cenário dos dados abertos de pesquisa. Assim, o CKAN novamente surge como um ferramental adequado para auxiliar os bibliotecários nesta tarefa, especialmente nas questões relativas a organização e a representação das informações. Na sequência demonstra-se como o CKAN pode contribuir para a execução destas funções no contexto dos dados abertos.

Primeiramente, é necessário que os dados estejam organizados atendendo as políticas institucionais no que tange as estruturas de organização de departamentos. Para tal, o CKAN fornece meios para que os bibliotecários possam estruturar as informações respeitando estas diretrizes. Como relatado anteriormente, a estrutura base para organizar os dados no sistema são as “Organizações”, em que, necessariamente, todo conjunto de dados deve estar vinculado a uma organização. Neste sentido, no contexto das bibliotecas e dos dados de pesquisas, as organizações podem ser utilizadas para realizar uma classificação dos dados, como por exemplo, criando uma organização para cada departamento de uma universidade ou para cada programa de pós-graduação.

No entanto, o CKAN oferece uma segunda forma de estruturar os dados, que são os chamados “Grupos”. Os grupos funcionam como uma estrutura paralela às organizações, em que um *dataset* pertencente a uma organização pode estar vinculado a inúmeros grupos. Desta forma, os grupos podem ter uma função importante para melhorar o processo tanto de organização quanto de recuperação das informações, ao oferecer outras formas de classificar os dados.

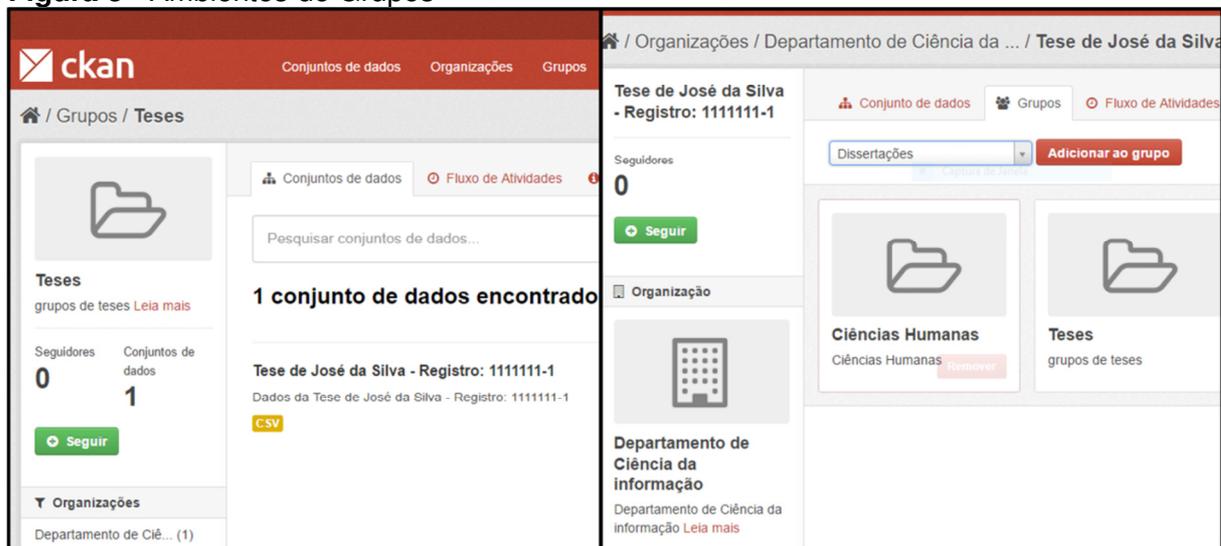
Para demonstrar como a estrutura de grupos pode auxiliar no trabalho do bibliotecário, realizou-se um teste em que os dados pertencentes a uma “Organização”, torna-se vinculado a grupos que classificam o conteúdo destes dados. Para tal, primeiramente foi criada a organização “Departamento de Ciência da Informação”, seguindo as recomendações dadas anteriormente no que se refere a criação de organizações e em seguida foi realizado a inserção de um conjunto de dados referentes a “Tese de José da Silva - Registro: 1111111-1”, que trazia os dados abertos de pesquisa de uma determinada tese, seguindo os passos descritos previamente, tratando da inserção de conjuntos de dados.

Na sequência, realizou-se a criação de grupos para que um conjunto de dados pudesse ser vinculados a eles, sendo inserido este conjunto de dados nos grupos referentes a “Teses” e a “Ciências Humanas”. O processo de criação de um grupo assemelha-se com o de criação de organizações, contudo apresentando objetivos distintos como relatado. Outra característica que distingue os grupos das organizações, é que a inclusão de um conjunto de dados só pode ocorrer por meio de uma organização, e somente depois pode ser realizada a vinculação deste conjunto com um grupo. Os grupos podem também ser utilizados como elementos

para a recuperação da informação, em que um usuário pode ir em um grupo para buscar por uma determinada informação.

A partir da criação dos grupos, é possível então, realizar a vinculação dos dados, acessando a aba de grupos em um conjunto de dados. A Figura 3 demonstra dois ambientes distintos referentes aos grupos, o ambiente da esquerda contém um grupo com um *dataset* vinculado, enquanto o da direita demonstra a aba “Grupos” de um conjunto de dados. Vale destacar que no primeiro ambiente demonstrado, não há a existência de uma opção para vincular dados, pois tal vinculação só pode ocorrer diretamente nos conjuntos.

Figura 3 - Ambientes de Grupos



Fonte: Elaborado pelos autores.

Além das estruturas de organização, o CKAN fornece elementos para a representação dos dados, como foi demonstrado anteriormente quando se discorreu sobre a representação realizada durante a inserção de conjuntos de dados. No entanto, no caso de dados de pesquisa há uma necessidade de uma representação mais refinada.

O CKAN possui um padrão de metadados para realizar a representação dos dados, sendo uma estrutura que descreve, ao menos parcialmente, o conjunto de dados, além de possuir três campos abertos em que o indivíduo no momento da descrição define tanto o nome do campo quanto o seu valor. Esses campos podem ser úteis e aprimorar a representação dos dados, uma vez que podem ser utilizados como um *link* para outro ambiente que contém elementos relacionados. Por exemplo, no caso de uma tese pode ser utilizado um campo para fazer um *link* ao

registro desta tese em repositório digital, contendo o objeto textual, enquanto no CKAN ficam organizados os dados referentes a pesquisa.

Porém, vale ressaltar que não há nas configurações básicas, opções para que seja utilizado outros padrões de metadados para realizar a representação dos dados, que pode ser visto como um problema, pois permite somente a utilização de seu próprio padrão de metadados, por mais que este padrão tenha uma relação com o padrão Dublin Core.

Uma outra característica que pode agregar na disponibilização dos dados de pesquisa no CKAN, é a possibilidade de inserção de diversos dados em um conjunto de dados. Essa possibilidade permite com que um conjunto de dados apresente dados distintos de uma mesma pesquisa, como por exemplo, os dados que foram gerados em uma entrevista, em que os dados brutos das respostas podem estar disponíveis em CSV e uma análise preliminar pode estar em PDF, além de permitir o acesso ao PDF do texto publicado

A partir das demonstrações e constatações realizadas, verifica-se que o uso do CKAN promove com que os dados de pesquisa fiquem organizados e estruturados, podendo fornecer aos usuários diversas informações que se relacionam, além de possibilitar a recuperação vindoura destas informações, contribuindo para a Ciência Aberta e para a Ciência como um todo. O bibliotecário encontra neste *software* um portal que aprimora a realização de suas funções, uma vez que tal ferramenta se mostra em consonância com as recomendações de dados abertos e possui manuseio simples.

5 CONSIDERAÇÕES FINAIS

A evolução das Tecnologias de Informação e Comunicação conduziram a um momento em que os dados são cada vez mais valorizados, sendo um dos principais elementos para a geração de valor para as organizações. Uma consequência desse processo foi a necessidade de abertura dos dados em alguns casos, como no âmbito governamental, para que a população pudesse ter acesso aos dados referentes ao estado e no cenário de dados de pesquisas, visando haver uma maior colaboração científica.

Neste contexto, o movimento de *Open Access* e de *Open Data* ganharam expressão, levando a definição de diretrizes e normas referentes a abertura de

dados. Além disso, diversas ferramentas foram desenvolvidas buscando facilitar este movimento, dentre essas ferramentas, destaca-se o CKAN, como uma plataforma para a disponibilização de dados abertos e para o *Linked Data*.

Assim, o profissional da informação apresenta um papel fundamental para a organização e a representação destes dados abertos, pois estes se tornaram um objeto informacional de extrema relevância e que necessita estar descritos e estruturados adequadamente para uma recuperação posterior. O CKAN por estar inserido no cenário dos dados abertos, pode auxiliar o profissional da informação na realização desta tarefa, sendo uma ferramenta completa que foi concebida de acordo com as necessidades deste domínio.

O presente trabalho demonstrou que o CKAN contém funcionalidades e é estruturado de forma que o profissional da informação, especialmente o arquivista e o bibliotecário, consigam implantar uma plataforma de dados abertos tanto no contexto de arquivos, tratando principalmente no meio governamental e empresarial, quanto no contexto de bibliotecas, especialmente tratando de dados abertos de pesquisas. As estruturas como “organização” e “grupos” se mostram como meios eficientes de disponibilizar os dados, permitindo uma classificação adequada, tornando o processo de recuperação eficaz.

No que diz respeito a forma como os dados são disponibilizados, o CKAN fornece meios que ao apresentar os dados fica explícito os formatos em que os dados estão disponíveis, além de possibilitar que sejam inseridos dados dos mais variados formatos. Tal questão favorece com que questões legais sejam atendidas, bem como permite aos usuários uma visão clara dos tipos de dados que estão disponíveis, não levando o usuário a confusões quanto aos formatos dos dados.

O CKAN fornece também meios para representar os dados por meio do uso de metadados baseados no Dublin Core. Entretanto, nos testes realizados não há a possibilidade de ser realizadas alterações quanto ao padrão utilizado, permitindo somente o uso do padrão próprio. Tal questão se mostrou como um problema, pois por vezes os arquivistas e os bibliotecários podem preferir utilizar outros padrões para estar em consonância com determinadas políticas ou outros serviços fornecidos.

De todo modo, o padrão do CKAN permite uma representação adequada, além de possuir campos em brancos, para que sejam inseridos outros metadados.

Vale destacar que o CKAN se mostra como uma ferramenta importante no contexto do *Linked Data* e Web Semântica, no que tange a disponibilização de conjuntos de dados. Tal questão demonstra a abrangência desta ferramenta, sendo uma outra possibilidade de atuação do profissional da informação.

Portanto, esta pesquisa demonstra que o CKAN, por mais que apresente alguns problemas, é um sistema que possibilita que arquivistas e bibliotecários desenvolvam suas tarefas no que concerne a dados abertos, principalmente questões relativas a organização e a representação das informações. O CKAN por ser um ambiente desenvolvido pela organização que trata justamente das questões dos dados abertos, se mostra como aderentes as normativas e as recomendações estabelecidas, sendo apropriado para os contextos de arquivos e bibliotecas e para os seus respectivos profissionais.

REFERÊNCIAS

- BARBOSA, R. R. Perspectivas profissionais e educacionais em biblioteconomia e ciência da informação. **Ci. Inf.**, Brasília, v. 27, n. 1, p. 53-60, jan./abr. 1998. Disponível em: <<http://www.scielo.br/pdf/ci/v27n1/07.pdf>>. Acesso em: 15 jun. 2016.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific american**, v. 284, n. 5, p. 28-37, 2001. Disponível em: <[https://kask.eti.pg.gda.pl/redmine/projects/sova/repository/revisions/master/entry/doc/Master%20Thesis%20\(In%20Polish\)/materials/10.1.1.115.9584.pdf](https://kask.eti.pg.gda.pl/redmine/projects/sova/repository/revisions/master/entry/doc/Master%20Thesis%20(In%20Polish)/materials/10.1.1.115.9584.pdf)>. Acesso em: 02 jun. 2016.
- BERNERS-LEE, T. **Linked data**. 2006. Disponível em: <<https://www.w3.org/DesignIssues/LinkedData.html>>; Acesso em: 03 jun. 2016.
- CKAN. **Overview**. 2013. Disponível em: <<http://docs.ckan.org/en/latest/>>. Acesso em: 02 jun. 2016.
- CKAN. **About**. 2015a. Disponível em: <<http://ckan.org/about/>>. Acesso em: 03 jun. 2016.
- CKAN. **Features**. 2015b. Disponível em: <<http://ckan.org/features/>>. Acesso em: 30 maio 2016.
- COSTA, M. M.; CUNHA, M. B. O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais. **Perspectivas em Ciência da Informação**, v.19, n.3, p.189-206, jul./set. 2014. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/download/1900/1445>>. Acesso em: 30 maio 2016.

DIETRICH, D. et al. **Open data handbook**. 2009. Disponível em:
<http://opendatahandbook.org/guide/pt_BR>. Acesso em: 11 jun. 2016.

LÉVY, P. **A inteligência coletiva: por uma antropologia do ciberespaço**. 10. ed. São Paulo: Edições Loyola, 2015.

OPEN KNOWLEDGE BRASIL. **Sobre**. 2016. Disponível em:
<<http://br.okfn.org/sobre/>>. Acesso em: 12 jun. 2016.

RIBEIRO, F. O perfil profissional do arquivista na sociedade da informação. **Trabalhos de Antropologia e Etnologia**, v.45, n.1, p.49-57, 2005. Disponível em:
<<https://repositorio-aberto.up.pt/bitstream/10216/14077/2/fribeiroperfil000108395.pdf>>. Acesso em: 18 jun. 2016.

RODRIGUES, F. A.; SANT'ANA, R. C. G. Restrições tecnológicas e de acesso a dados disponíveis sobre destinos de repasses financeiros federais para a saúde pública em ambientes informacionais digitais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012. **Anais...** Rio de Janeiro: FIOCRUZ, 2012. Disponível em:
<<http://www.eventosecongressos.com.br/metodo/enancib2012/arearestrita/pdfs/19435.pdf>>. Acesso em: 10 jun. 2016.

SANTAREM SEGUNDO, J. E. Web Semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente as iniciativas internacionais. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v.8, n.2, p.219-239, jul./dez. 2015. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/view/207>>. Acesso em: 16 jun. 2016.

SANTAREM SEGUNDO, J. E.; FARIA, T. EXPLORANDO TÉCNICAS E RECURSOS DO GERENCIADOR DE DADOS ABERTOS CKAN. In: ENCONTRO INTERNACIONAL DADOS, TECNOLOGIA E INFORMAÇÃO, 2013, Marília/SP, **Anais eletrônicos...** Disponível em:
<<http://gpnti.marilia.unesp.br:8085/index.php/DTI/DTI/paper/viewFile/285/106>>. Acesso em: 20 jun. 2016.

SANTAREM SEGUNDO, J. E. Tecnologias de la informacion y la comunicacion para proporcionar datos abiertos en formato semántico. **Ibersid**, v. 7, p. 33-40, 2013. Disponível: <<http://ibersid.eu/ojs/index.php/ibersid/article/view/4075>>. Acesso em: 15 jun. 2016.

SAYÃO, L. F.; SALES, L. F. Dados abertos de pesquisa: ampliando os conceitos de acesso livre. **RECIIS**, v. 8, n. 2, p. 76-92, 2014. Disponível em:
<https://www.researchgate.net/profile/Luana_Sales2/publication/263618733_Dados_abertos_de_pesquisa_ampliando_o_conceito_de_acesso_livre/links/00b7d53b5edad560fa000000.pdf>. Acesso em: 13 jun. 2016.

WINN, J. et al. **Open data and the academy**: An evaluation of CKAN for research data management. 2013. Disponível em:
<<http://eprints.lincoln.ac.uk/9778/1/CKANEvaluation.pdf>>. Acesso em: 04 jun. 2016.

Title

Information Professional in the context of Open Data: the use of CKAN for the availability and organization of data

Abstract

Introduction: The development of technologies is promoting an increase in data generation. In this sense, there are several areas that need to promote the opening of their information, featured companies, institutions, governments and academia. The information professional may have a role in this scenario, though it is necessary that technologies are used which contribute to this, as CKAN tool, which promotes the availability of open data.

Objective: To discuss the use of CKAN by archivists and librarians for the organization and representation of data in open formats.

Methodology: Exploratory qualitative, with theoretical knowledge about open data, the CKAN system and information professionals, and theoretical applied as discussed using CKAN system for the roles of information professionals, performing tests on a system in place to this paper.

Results: Through theoretical discussions and tests within the tool, we demonstrate the CKAN's relation to the activities of information professionals, and how such a tool can be used to perform these activities.

Conclusions: CKAN shown as a suitable system for the organization and representation of data open for both archivists and for librarians by providing features that facilitate the publication of open data.

Keywords: Open Data. Information Profesional. Archivist. Librarian. CKAN.

Titulo

Profesional de la información en el contexto de datos abiertos: el uso de CKAN para la disponibilidad y organización de datos

Resumen

Introducción: El desarrollo de las tecnologías está promoviendo un aumento en la generación de los datos. En este sentido, existendiversas áreas que necesitan promover la apertura de sus informaciones, tales como empresarial, institucional, gubernamental y académica. El profesional de la información puede tener un papel en este escenario, siendo necesario que se utilicen las tecnologías que contribuyan para esto, como la herramienta CKAN, que promueve la disponibilidad de datos abiertos.

Objetivo: Discutir el uso de CKAN por los archivistas y bibliotecarios para la organización y representación de los datos en formatos abiertos.

Metodología: cualitativa exploratoria, con base teórica sobre los datos abiertos, los profesionales de la información y el sistema CKAN y teóricos aplicados como se discute el uso del sistema de CKAN para los papeles de profesionales de la información, realizando pruebas en un sistema implantado para este trabajo.

Resultados: A través de discusiones teóricas y pruebas dentro de la herramienta, se demuestra la relación del CKAN con las actividades de los profesionales de la información, y cómo una herramienta de este tipo puede ser utilizada para llevar a cabo estas actividades.

Conclusiones: CKAN se muestra como un sistema adecuado para la organización y representación de datos abiertos tanto para los archivistas y los bibliotecarios, proporcionando características que facilitan la publicación de los datos abiertos.

Palabras clave: Datos abiertos; Profesional de la información; Archivero; Bibliotecario; CKAN

Recebido em: 25.08.2016

Aceito em: 30.11.2016