

# Estratégias para Abertura de Dados da Universidade Federal do Rio Grande do Norte

Andressa Kroeff Pires<sup>1</sup>, Gibeon Soares de Aquino Júnior<sup>1</sup>,  
Gusttavo Henrique N. Sousa Silva<sup>1</sup>, Jean Guerethes F. Guedes<sup>1</sup>,  
Vicente Fidelis Ferreira Gomes Neto<sup>1</sup>

<sup>1</sup>Superintendência de Informática  
Universidade Federal do Rio Grande do Norte (UFRN)  
Caixa Postal 59078-970 – Natal – RN – Brazil

***Resumo.** Ao catalogar os dados abertos, é necessário garantir que os dados publicados estejam sempre atualizados, sejam disponibilizados em diversos formatos de arquivos que possam ser interpretados por pessoas ou máquinas, estejam organizados e classificados para que, assim, sejam facilmente encontrados. Para tanto, é necessário uma plataforma que garanta o cumprimento dessas e outras exigências. Neste artigo, será discutido sobre a plataforma CKAN como uma solução para a publicação, gerenciamento e consumo de dados em instituições de caráter governamental.*

## 1. Introdução

Os órgãos públicos geram grandes quantidades de dados, que podem ser transformados em informações úteis tanto para essas instituições, como para a sociedade, de modo amplo. A publicação desses dados, se usados para esse propósito, pode propiciar mais controle social, transparência, cidadania ativa, cooperação, melhorias na administração pública, bem como a criação de novas ferramentas que irão garantir mais desenvolvimento e bem-estar social [Kucera 2015].

No Brasil, várias instituições do governo já disponibilizam seus dados. Porém, o número de instituições ainda é pequeno em relação à outros governos (como o dos E.U.A., Reino Unido, Canadá e Nova Zelândia). Em novembro de 2011 foi instituída a lei de acesso à informação (LAI), que regulamenta o direito constitucional de acesso às informações públicas. A partir dessa lei foram criados mecanismos que possibilitam a qualquer pessoa, física ou jurídica, sem necessidade de apresentar motivo, o recebimento de informações públicas dos órgãos e entidades do governo [Ribeiro 2012]. Devido à essa e outras medidas adotadas pelo governo federal, as instituições públicas passaram a se preocupar em divulgar seus dados, por isso a tendência é de que o número de instituições cresça em um curto intervalo de tempo.

Embora a filosofia do acesso à informação seja conhecido há alguns anos, há pouco tempo o conceito de dados abertos tornou-se conhecida devido à algumas iniciativas adotadas por governos de alguns países para a abertura de suas informações públicas. Na definição da *Open Knowledge*, dados abertos são dados que podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa – sujeitos, no máximo, à exigência de atribuição à fonte original e ao compartilhamento pelas mesmas licenças em que as informações foram apresentadas [sit b]. Os dados abertos não se restringem à documentos textuais, mas se estendem por uma série de formatos, como: imagens,

planilhas, *Linked Data* ou qualquer outra mídia que possa ser transformada em novas informações [sit a].

Nessa perspectiva, é possível vislumbrar os produtos obtidos a partir da abertura dos dados. Porém, para que esses produtos sejam possíveis é preciso garantir a continuidade na prática da publicação dos dados com o comprometimento das instituições no compartilhamento de seus dados públicos e seu empenho na atualização periódica dos mesmos. É necessário, também, que haja uma infraestrutura que seja capaz de armazenar a massiva quantidade de dados que serão disponibilizados e que atenda a grande quantidade de acessos por parte dos usuários [JUNIOR 2016].

Para isso, faz-se necessário estudar estratégias para o desenvolvimento de uma infraestrutura que viabilize a publicação, atualização e consumo dos dados. O objetivo deste artigo é apresentar as estratégias adotadas pela Superintendência de Informática da Universidade Federal do Rio Grande do Norte (UFRN) para a construção de um portal de catalogação dos dados abertos desta instituição. Nas próximas sessões serão explicitadas as estratégias para o desenvolvimento do portal, para a escolha desses dados, para sua publicação e atualização, bem como os problemas enfrentados e os resultados alcançados.

## 2. Métodos

Após um aprofundado estudo sobre os meios de compartilhamento de dados existentes (sites, API, ou até mesmo através de servidores de Protocolo de Transferência de Arquivos), decidiu-se por desenvolver o portal de dados da UFRN com a plataforma CKAN (*Comprehensive Knowledge Archive Network*) que foi desenvolvida pela OFK (*Open Knowledge Foundation*) e é a plataforma mais utilizada dentre as diversas organizações internacionais e nacionais que abriram seus dados.

Escolheu-se pelo CKAN, principalmente, por ser de código aberto, com uma ampla comunidade, inteiramente documentado, altamente personalizável, com fácil implantação e de fácil uso. O CKAN oferece uma intuitiva interface WEB e uma API (*Application Programming Interface*) que facilita o gerenciamento e consumo de seus dados e metadados de forma manual ou automatizada (através de sistemas computacionais externos).

O portal de dados da UFRN foi desenvolvido com o CKAN na versão 2.6.0 e sua arquitetura apresenta a seguinte configuração de máquinas:

- **UFRN-web** - Responsável pelo armazenamento dos arquivos de configuração do httpd e nginx. Contém o código open source.
- **UFRN-solr** - Reservada para o indexador de páginas do frontend, contém o tomcat.
- **UFRN-db** - Reservada para o banco de dados da aplicação, contém o banco postgresql.

Esta configuração foi adotada, seguindo as instruções da OFK, para garantir melhor desempenho da aplicação.

Uma das características do CKAN é que ele permite a instalação de plugins da própria plataforma, de terceiros ou desenvolvidos independentemente (com fins específicos) para complementar a ferramenta. No portal de dados abertos da UFRN, foram mantidos os plugins já pré-instalados e foram instalados/configurados plugins para a

visualização de dados em diversos formatos, gerenciamento de dados e visualização de estatísticas do portal. Abaixo estão alguns dos plugins utilizados no catálogo da UFRN:

- **Filestore** - para armazenamento dos arquivos;
- **Dataproxy** - serviço web para exibição de recursos de dados em forma estruturada, como JSON, CSV, XLS etc.;
- **Datapusher** - para extrair os dados dos arquivos e inserir no datastore;
- **Datastore** - para armazenamento de dados contidos em arquivos;
- **Multilingual** - para tradução da plataforma em outras línguas;
- **DataStore Grid, DataStore Map, DataStore Graph, Text View, Image View, Web Page View** - para a visualização personalizada dos dados;
- **Stats Extension** - extensão para estatísticas do CKAN.

Outra forte característica do CKAN é que ele permite a customização do *layout* do portal. É possível escolher entre os 3 templates oferecidos e ainda alterar seu esquema de cores e imagens. Ao término da instalação da plataforma foi iniciado um processo de customização das páginas do portal que não atendia aos padrões de *layout* da instituição.

Ao término do processo de customização, fora iniciado os estudos para decidir sobre quais dados seriam publicados e a frequência com que eles seriam atualizados. Este processo foi feito juntamente com uma comissão formada pela instituição para garantir que os dados publicados fossem dados de caráter público.

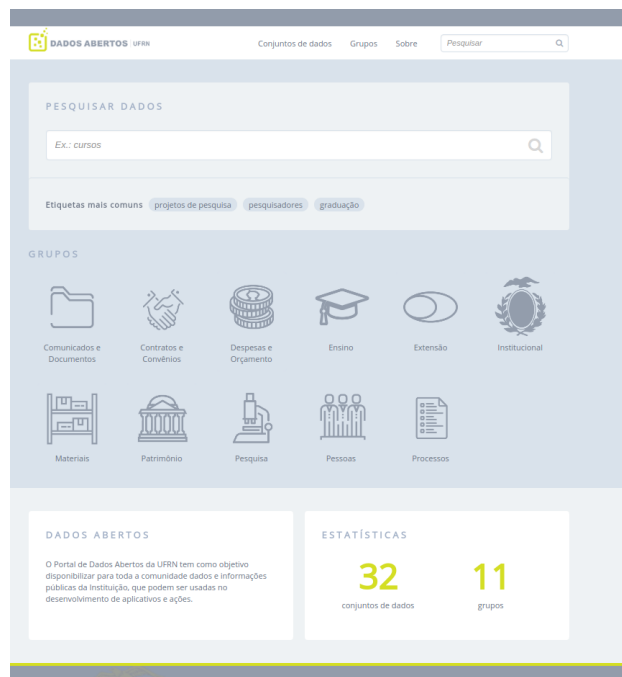
Para garantir a atualização frequente dos dados publicados, já que alguns dados apresentam alto grau de variação em um pequeno espaço temporal, foi desenvolvida uma ferramenta que realiza essa tarefa de forma automática. Na ferramenta, basta que se defina o datasource onde serão consultados os dados, a consulta que irá gerar o arquivo e o tempo de atualização do conjunto de dados. De tempos em tempos, a ferramenta busca pelos dados desatualizados, gera um novo arquivo daquele conjunto de dados e realiza o *upload* do mesmo utilizando a API do CKAN. Dessa forma é garantido que os dados publicados na portal estejam sempre o mais atualizado possível.

### 3. Resultados

Depois de finalizados os processos de instalação, configuração e customização do *layout*, criou-se as organizações e grupos e publicou-se os primeiros conjuntos de dados. A Figura 1 mostra o estado atual do portal de dados abertos da UFRN, que atualmente contém a organização UFRN, 11 diferentes grupos e 32 diferentes conjuntos de dados, cada conjunto de dado apresenta um ou mais arquivos brutos (não tratado) e um dicionário de dados para que o usuário final possa entender o que representa cada um dos atributos apresentados.

Na área central da página inicial, do portal de dados abertos, estão sendo exibidos todos os grupos de dados existentes, a barra de pesquisa foi mantida da versão original e um view do plugin de estatística dos dados contidos no portal foi adicionado no canto superior direito, para que o usuário tivesse uma visão geral do quantitativo de dados existentes no portal.

Foi feito um levantamento dos acessos ao portal e foi observado que diariamente cerca de 60 pessoas acessam e consomem os dados publicados no portal. O uso do

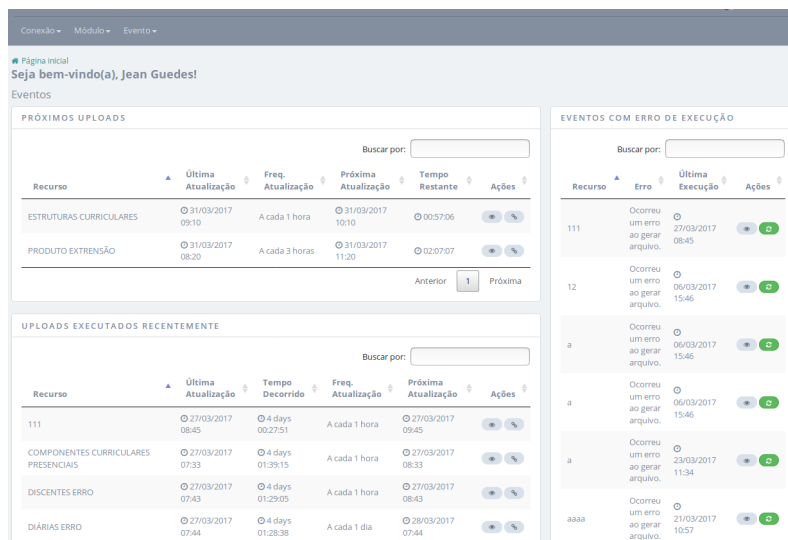


**Figura 1. Layout final do portal de dados abertos da UFRN.**

catálogo também se confirma através dos emails recebidos pela equipe de desenvolvimento sugerindo a criação de novos conjuntos de dados.

A ferramenta criada para a atualização automática dos dados ainda se encontra em homologação, já apresentando bons resultados na atualização dos dados publicados no portal.

Na Figura 2 é possível ver o *dashboard* da plataforma indicando os dados que foram atualizados recentemente, as próximas atualizações e as atualizações que apresentaram erro ao gerar o arquivo ou ao fazer o *upload* para o portal.



**Figura 2. Sistema de atualização dos dados abertos.**

#### 4. Conclusão

O CKAN, por ser um ambiente desenvolvido pela organização que trata justamente das questões dos dados abertos, mostra-se aderente às normativas e as recomendações estabelecidas, sendo apropriado para o contexto de dados abertos. Além de tudo, mostra-se como uma ferramenta importante no contexto do *Linked Data* e Web Semântica, no que tange a disponibilização de conjuntos de dados.

No que diz respeito à forma como os dados são disponibilizados, o CKAN fornece meios que, ao apresentar os dados, ficam explícitos os formatos em que os dados estão disponíveis, além de possibilitar que sejam inseridos dados dos mais variados formatos. Tal questão favorece para que questões legais sejam atendidas, bem como permitir aos usuários uma visão clara dos tipos de dados que estão disponíveis, não levando o usuário a confusões quanto aos formatos dos dados. Além disso, as estruturas como “organização”, “grupos” e “*datasets*” se mostram como meios eficientes de disponibilizar os dados, permitindo uma classificação adequada e tornando o processo de recuperação eficaz.

Portanto, por mais que o CKAN ainda apresente alguns problemas, o sistema se mostra capaz de atender às demandas de publicação de dados de órgãos do âmbito público e privado devido sua abrangência, sendo uma forte candidato para implementação de mecanismos de compartilhamento de dados. A plataforma pode ser implantada para atender as necessidades de publicação de dados abertos em diversos contextos nas áreas da ciência, educação, saúde, transporte, meio ambiente e outros.

#### Referências

- Linked open data: The essentials. a quick start guide for decision makers. <http://5stardata.info>. Acesso: 05-12-2016.
- Open knowledge foundation: The open data handbook (2012). <http://opendatahandbook.org>. Acesso: 24-11-2016.
- JUNIOR, GIBEON S A e SOUSA S, G. H. N. e. S. R. O. A. (2016). Estratégia para desenvolvimento do portal de dados abertos da cidade de natal. *III WORKSHOP SOBRE CIDADES INTELIGENTES*, pages 1–2.
- Kucera, Jan e Chlapek, D. e. K. J. e. N. M. (2015). Methodologies and best practices for open data publication. In *DATESO*, pages 52–64.
- Ribeiro, C e Almeida, R. d. F. (2012). Towards open government data: The publication of the brazilian social security systems statistical data. In *The 6th International Multi-Conference on Society, Cybernetics and Informatics IMSCI. Orlando, Florida*.