

Uso de software livre para disseminação e análise de dados abertos governamentais

Lucas Rodrigues Costa

Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT, E-mail: lucasrodrigues@ibict.br

Lucas Ângelo Silveira

Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT, E-mail: lucasangelo@ibict.br

Rennie Fagundes Brito

Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT, E-mail: ronniebrito@ibict.br

Milton Shintaku

Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT, E-mail: shintaku@ibict.br

RESUMO

O estudo apresenta um modelo voltado à disseminação e análise de dados de órgãos governamentais com softwares livres. Foi adotada uma metodologia voltada a seleção de tecnologias robustas que atendessem às demandas das agências envolvidas. O modelo baseou-se nos softwares Comprehensive Knowledge Archive Network (CKAN) e Pentaho, os quais possibilitam serviços de depósito, recuperação, visualização e análise dos dados. O caso de estudo foi aplicado na Secretaria Nacional da Juventude, onde dados oriundos de várias fontes foram recolhidos, tratados e publicados para o público em geral. O modelo vai de encontro à objetivos dos órgãos governamentais, como a oferta de acesso a dados brutos e sua análise pelo público em geral, contribuindo para com a transparência de seus respectivos dados.

Palavras-chave: Softwares livres. Dados abertos. Data warehouse.

1 INTRODUÇÃO

É indiscutível que a internet tem atuado em todo o mundo na disseminação da informação e oferta de serviços em todas as áreas, inclusive nos governos. Com isso, altera-se o comportamento dos cidadão, que cada vez mais requerem serviços e informações online, de forma a facilitar e agilizar as suas necessidades. Além disso, torna-se possível acompanhar as ações do governo, como forma de ampliar a cidadania.

No Brasil a transparência do estado tem sido promovida com várias ações, entre os quais situa-se a iniciativa para dados abertos governamentais, processo pelo qual os governantes disponibilizam informações aos seus cidadãos (OLIVERIO, 2011). Essa orientação governamental é regida pela Lei nº 12.527, que regula o acesso livre à informação

governamentais, reservadas às questões de segurança e proteção dos dados sensíveis. Essa lei engloba desde dados brutos à documentos completos, tratando questões como sigilo, autenticidade, integridade e primariedade, entre outros pontos (BRASIL, 2011).

Com isso, o governo espera promover a disseminação ampla das informações governamentais, na medida em que são consideradas públicas e vistas por Jardim e Marcondes (2003) como um dos principais componentes de mediação democrática. Assim, suas especificidades dependem do contexto a que é criada assim como os seus formatos e tipos. No presente estudo, estuda-se os dados governamentais na perspectiva de dados abertos, como forma do governo promover a transparência.

Em uma análise detalhada, dados governamentais são descritos como resultados de atividades dos órgãos públicos e podem estar contidos em bases de dados, documentos impressos ou digitais, entre outros (OLIVEIRA, 2016), figurando assim uma grande variedade de formas e formatos. Por sua vez, dados abertos podem ser lidos, utilizados e disseminados de forma livre, sendo requisitos a citação da fonte e se for o caso o compartilhando sob mesmo tipo de licença (ISOTANI; BITTERN COURT, 2015). Assim, dados abertos governamentais são dados abertos gerados pelo governo e que devem ser disseminados com a sociedade, podendo ser reutilizados para análises diversas, acadêmicas ou estratégicas, por exemplo.

Palazzi e Tygel (2014) afirmam que parte desses dados gerados no governo são de cunho estatístico, possibilitando a geração de indicadores, que podem ser utilizados para análises e tomada de decisão. Dessa forma, para a disseminação e análise dos dados governamentais, pode-se utilizar sistemas informatizados que possibilitem a implementação de políticas relacionadas iniciativa de dados abertos governamentais para a disseminação e análise dos dados armazenados e produzidos do governo.

Com isso, tem-se oportunidades e desafios relacionados à pesquisas que atendam às necessidades das instituições e órgãos às suas especificidades ao tratamento dos dados abertos governamentais, de forma a atender com maior eficácia aos anseios da população. Assim, o presente trabalho apresenta o resultado de pesquisa efetuado na Secretaria Nacional de Juventude (SNJ), voltado à criação de um modelo de dados abertos governamentais utilizando software livre. Isso, contribui com a discussão sobre dados abertos governamentais e a disseminação de informação por meio de software livre.

2 PROCEDIMENTOS METODOLÓGICOS

O presente estudo tem aspectos aplicados, com utilização de técnicas ligadas à ciência da computação, na seleção e uso de tecnologias, para criação e aplicação de um modelo voltado a dados abertos com o uso de software livre. Assim, utiliza-se conceitos e técnicas voltados à avaliação de ferramentas informatizadas, alinhado à técnicas ligadas a qualidade de software. Nesse sentido, qualidade é entendida na forma da Norma ISO 8402, na qual se refere ao conjunto de características relacionadas ao atendimento das necessidades dos usuários, sejam explícitas ou implícitas. Atendendo os requisitos registrados ou não, a qualidade do software é uma avaliação de aspectos quantitativos e qualitativos, no qual contempla o processo e produto, como advogado por Tsukumo et al (1997), na medida em que o processo oferta certas garantias ao produto.

Para Silva (2007), a avaliação de softwares livres não deve ser embasada apenas na gratuidade da ferramenta, mas nos benefícios gerais que a nova tecnologia pode trazer. Seguindo tal raciocínio, o critério básico de seleção de ferramentas a serem utilizadas, era ser software livre e que provesse funcionalidades tais como: suporte por comunidade internacional atuante, mantida por instituição confiável, e que fornecesse certas garantias de sustentabilidade.

O estudo não foca apenas na avaliação do software como produto, mas também como serviço, na medida em que essas ferramentas devem disseminar dados de governo à população, alinhado com os preceitos de dados aberto de governo. Com isso, testa-se o uso da ferramenta como forma de avaliação, avaliando o seu desempenho em ofertar serviços de disseminação da informação de forma eficaz, com organização e representação adequadas. Ferramentas para disseminação e análise de dados.

O processo de seleção do software baseou-se nos critério de ser uma ferramenta livre, de código aberto, mantido por instituição reconhecida e apoiada por uma comunidade internacional ativa. Com isso, procura-se ofertar garantia maior de continuidade, na medida em que há um grande número de ferramentas livres que são descontinuadas.

A seleção adequada das ferramentas utilizadas para a disseminação e análise dos dados do governo tem o intuito de auxiliar a iniciativa de dados abertos governamentais com o propósito de divulgar dados e informação. Assim, garante-se não apenas a implementação de sistemas de informação para dados de governo, mas principalmente a sua continuidade.

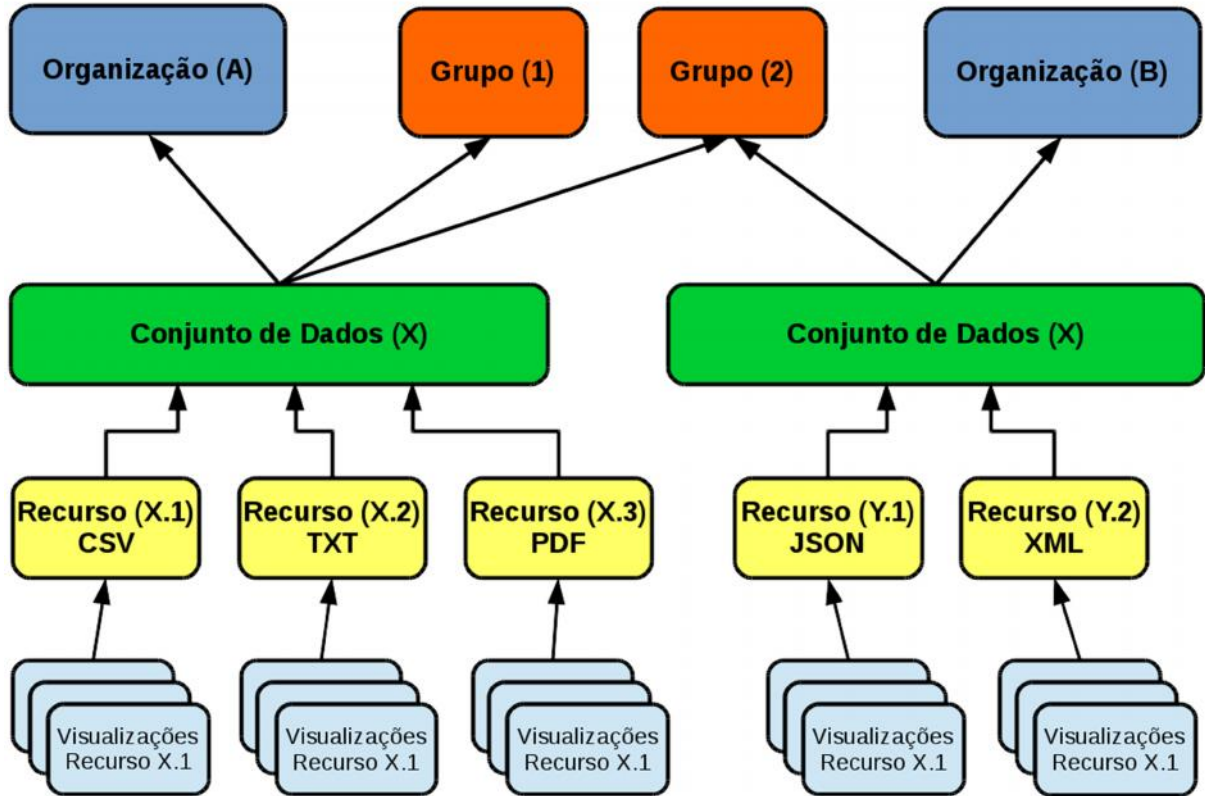
Seguindo os critérios estabelecidos para a seleção de software livres para o projeto e os benefícios fornecidos, selecionou-se as ferramentas para repositórios de dados Comprehensive Knowledge Archive Network - CKAN e o Pentaho.

A composição de tais ferramentas tem por objetivo utilizar o CKAN como um coletor e depósito de dados brutos que em sua essência organiza tais dados de forma coerente e concisa, num segundo estágio se aplica o Pentaho como uma ferramenta de mineração de dados para reutilizar e redistribuir os dados mantidos no CKAN oferecendo suporte à gestão e análise dos dados de acordo com o interesse da instituição. Com isso, atende-se às atividades de depósito e disseminação de bases de dados de governo pelo CKAN e o refinamento dos dados brutos, gerando informação através do Pentaho.

O Software CKAN

O CKAN (CKAN, 2017) possibilita o depósito de bases de dados de forma organizada. Conforme a hierarquia descrita por Costa et al (2017) na figura 1, tem-se os conjuntos de dados como ponto central ligados às organizações. O conjunto de dados é caracterizado e descrito por metadados, podendo conter um ou mais recursos, em formato diversos. Com isso, caso tenha-se uma base de dados com arquivos em formato textuais, planilhas entre outros. Estes dados podem ser agrupados em uma única base de dados ligados a uma organização, tendo em vista que um único CKAN pode ser utilizado por uma ou mais instituições. Além disso, pode-se criar organizações artificiais para agrupar uma ou mais bases de dados denotadas como grupos e com isso facilitar a recuperação dos dados.

Figura 1 - Possíveis hierarquias de base de dados no CKAN



Fonte: CKAN (2017).

Pode-se também destacar que o CKAN colabora de forma significativa a questão dos dados abertos e dos *Linked Data*, como descrevem Coneglian e Segundo (2017), na medida em que está preparado para disponibilizar bases de dados no formato do *Liked Open Data* (LOD). Neste cenário, promove o uso de formato de dados com aspectos semânticos, fomentando a recuperação dessas bases de dados.

Este modelo alinha-se às indicações de Correa et al. (2017) as quais descrevem o uso do CKAN como uma ferramenta que apoia a disseminação de dados governamentais, contribuindo em parte com a transparência do governo. Cabe ressaltar que o CKAN está de acordo com as orientações dos dados abertos governamentais e a Lei nº 12.527, que regula o acesso livre à informação de governo, de acordo com o Plano de Dados Abertos-PDA (<http://bit.ly/2vnCiWD>).

Destaca-se que, conforme o PDA, o CKAN é a ferramenta adotada pelo Ministério do Planejamento, Desenvolvimento e Gestão (MPOG), no desenvolvimento do Portal de Dados de

Abertos. Com isso, todos os sistemas de informações implementados com o CKAN se integram de forma automática, em forma de rede, compartilhando metadados.

Nesse contexto, pode-se verificar que instalações de portais de dados de órgãos governamentais desenvolvidos com o CKAN tornam-se elementos de redes de dados do Portal de Dados abertos do governo brasileiro. Com isso, contribui-se com a transparência, na medida em que cada órgão tem a independência sobre as suas bases de dados, garantindo o acesso e ao mesmo tempo participando de uma rede nacional.

O Software Pentaho

No que se refere a análise dos dados, optou-se pela ferramenta Pentaho, visto que é uma ferramenta robusta para *Business Intelligence* (BI), que apoia o tratamento de dados (AHISHAKIYE, 2017), com foco no tratamento e análise de dados (COSTA et al., 2017; MARINHEIRO; BERNARDINO, 2015). Assim, tem-se uma ferramenta livre que implementa funcionalidades de BI voltadas a análise de dados, podendo gerar indicadores que apoiem a tomada de decisão.

O Pentaho oferece funcionalidades relacionadas à *Online Analytical Processing* (OLAP) que possibilitam analisar grandes volumes de dados de forma online e sob diferentes cruzamentos e dimensões dos dados, bem como a realização de cálculos complexos. O principal produto envolvido em uma análise por meio de OLAP é o cubo multidimensional. Vale ressaltar, que a ferramenta é uma solução composta por vários módulos a fim de atender diferentes etapas de estruturação de uma base analítica. Entre os módulos disponíveis no Pentaho está o *Pentaho Data Integration* (PDI), o *Pentaho Schema Workbench* (PSW- MONDRIAN) e o SAIKU.

- a) PDI ou Kettle - tem por finalidade a integração de dados, possibilitando cruzar dados proveniente de várias fontes, com o uso de técnicas de ETL (*extract-transform-load*). O PDI oferece uma interface gráfica para a conexão das várias fontes e possibilita apresentar os resultados do processamento em formas de grafos. Para Casters, Bouman e van Dougen (2010) o Kettle é um mecanismo que resolve várias atividades, de forma eficaz, mesmo na forma padrão de uso.
- b) PSW - MONDRIAN - é uma ferramenta para o desenvolvimento de um esquema xml que descreve o cubo multidimensional dos dados em tempo real, visto por Afonso (2009)

como a materialização total ou parcial do desempenho num sistema de processamento analítico. Com isso, os cubos multidimensionais fornecem informações analíticas em tempo real referentes a um conjunto de dados.

- c) SAIKU - é um módulo para visualização dos dados do cubo de uma forma amigável e dinâmica. Para Lamas (2013), o Saiku é um navegador OLAP que atende aos usuários por ser leve e intuitivo, possibilitando executar muitos processamentos com as informações tratadas pelo Pentaho de forma simples.

Sallam et al (2014) relatam que o Pentaho se apresenta como uma solução para diversas atividades de BI, tendo uma grande presença no mercado. Assim, por ser uma ferramenta livre de pagamento de licença, torna-se uma opção válida e sustentável, alinhado aos preceitos governamentais de uso de software livre.

3 RESULTADOS

O presente estudo relata a implantação e os resultados do uso do CKAN e Pentaho em projeto de pesquisa desenvolvida na Secretaria Nacional de Juventude, de forma a atender as necessidades de alinhamento às orientações de dados abertos de governo e do acesso à informação. Assim, os resultados apresentados nesta seção são restritas a esse cenário, no entanto, acredita-se que possam se reproduzido por outras implementações.

A Secretaria Nacional da juventude (SNJ) possui um fluxo de dados no qual coleta ou gera uma grande quantidade de dados brutos sobre juventude, nos mais diversos temas correlatos. Conforme descreve Cury (2007), essa secretaria nasceu de uma ação interministerial, devido ao seu caráter multifacetado, envolvendo temas distintos que requerem atenção transversal ao governo. O mesmo autor, relata que o tema juventude é novo na política no mundo, revelando a inovação desta secretaria e por consequência suas ações.

Após o levantamento de requisitos, verificou-se que a principal necessidade da SNJ era de um modelo que contemplasse a gestão de bases de dados estatísticas, com o fluxo informacional de Coleta ou Geração; Catalogação; e finalmente sua Recuperação/Análise. Os dados em sua maioria, são provenientes de outras instituições como o Instituto Brasileiro de Geografia e Estatística (IBGE) e Instituto de Pesquisa Estatística Aplicada (IPEA), ou gerado para o SNJ, por instituições como a Caixa Econômica Federal. Assim no primeiro caso considera-se como coleta

e no segundo uma geração. Requerendo, dessa forma, uma ferramenta que possibilite o armazenamento organizado das bases de dados, permitindo a recuperação e análise dos dados.

A prospecção focou em duas etapas, sendo que a primeira consistiu na migração dos dados de seus diferentes formatos para uma plataforma comum, e uma segunda que trata de disponibilizar uma ferramenta de visualização de dados e elaboração de relatórios. Pela prospecção de tecnologias foi utilizado o uso integrado do CKAN e do Pentaho, na medida em que atendia os principais critérios definidos pelo projeto.

Com isso, tem-se o CKAN como um repositório de dados, com todas as funcionalidades voltadas à gestão de bases de dados e o Pentaho como uma ferramenta com foco em analisar, organizar e apresentar tais dados. Assim, atende-se as principais reivindicações da SNJ em atender ao dados abertos de governo e o acesso à informação.

CKAN na SNJ

De forma padrão o CKAN está preparado para gerenciar dados de várias organizações, como um repositório agregador, no qual a classificação das bases de dados não tem o viés temático. Assim, foi mister ajustar as características da ferramenta às necessidades de disseminação organizadas tematicamente das bases de dados da SNJ, na medida em que essas bases de dados possuem características temáticas.

Na SNJ as organizações artificiais (grupos) do CKAN tornaram-se temas de interesse da secretaria, como: saúde, educação, lazer, onde as organizações são organizadas como sendo diferentes fontes para base de dados. Com isso, tem-se um modelo de repositório para dados governamentais, organizado de forma temática, com três níveis, sendo: Tema base de dados recursos, possibilitando catalogar todas as bases de dados de forma organizada, facilitando a recuperação.

A figura 2 apresenta a organização das bases de dados do Portal de dados da SNJ, no qual os temas são implementadas com as organizações artificiais (grupos). Cada grupo representa um tema tratado pela SNJ, destacando a transversalidade das ações e dados mantidos pela secretaria, de forma que gestores e usuários em geral consigam facilmente recuperar as bases de dados mantidas pelo portal.

Figura 2 - Página do CKAN do SNJ.



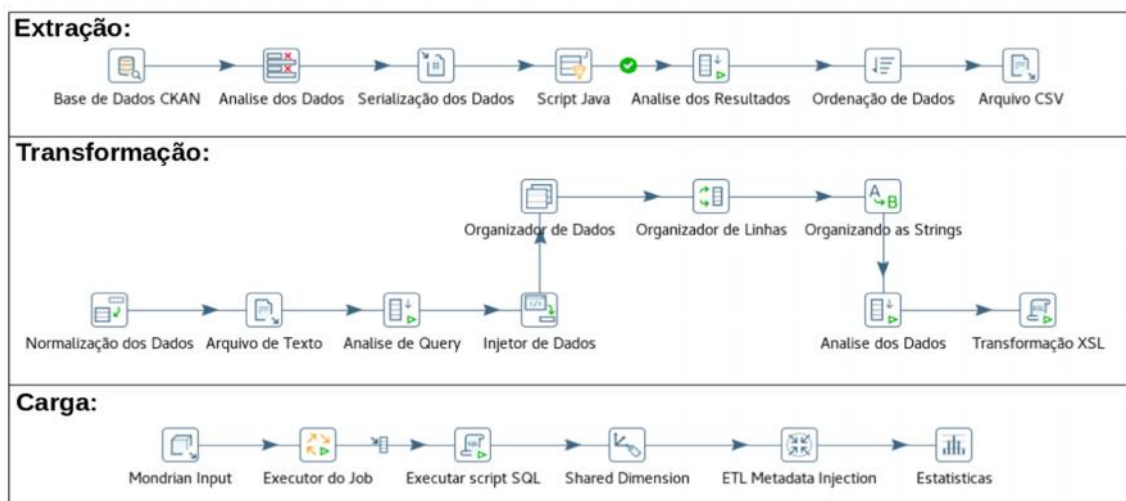
Fonte: Elaborada pelos autores (2017).

O CKAN neste modelo apresenta-se como repositório de dados tornando-se útil para a catalogação das bases de dados em uma estrutura organizada e integrável com outros sistemas CKAN na forma de um ecossistema de dados governamentais. Todavia, por ter um único órgão depositante de dados, tornou-se um repositório de dados institucional e temático.

Pentaho na SNJ

Inicialmente foi desenvolvido uma estratégia para a extração dos dados armazenados no CKAN e uma transformação dos mesmos para um formato comum, tendo em vista que diferentes bases e tipos de dados foram armazenados no repositório. Para essa tarefa foi utilizado o módulo PDI que possibilitou cruzar dados proveniente de diversas fontes do CKAN por meio das técnicas de ETL. A interface gráfica para a conexão de tais fontes auxiliou na apresentação das tarefas resultando nos processos descritos na figura 3.

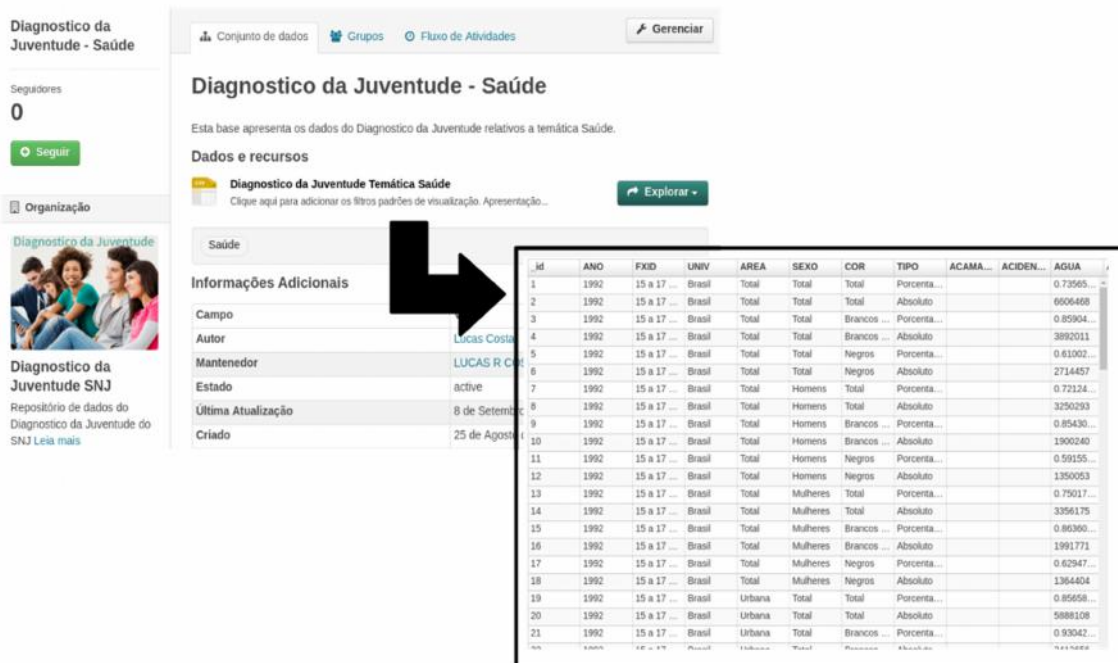
Figura 3 - ETL utilizado no processo da SNJ.



Fonte: Elaborada pelos autores (2017).

Como pode ser visto, foram separados as três etapas do processo ETL realizadas no PDI. A extração começa na base de dados do CKAN a qual possuem inúmeras tabelas com informações de diversas áreas relacionadas à juventude. A figura 4 mostra um exemplo do conjunto de dados relacionado a saúde da juventude no Brasil e como se encontra o formato dos dados.

Figura 4 - Exemplo de conjunto de dados da SNJ.



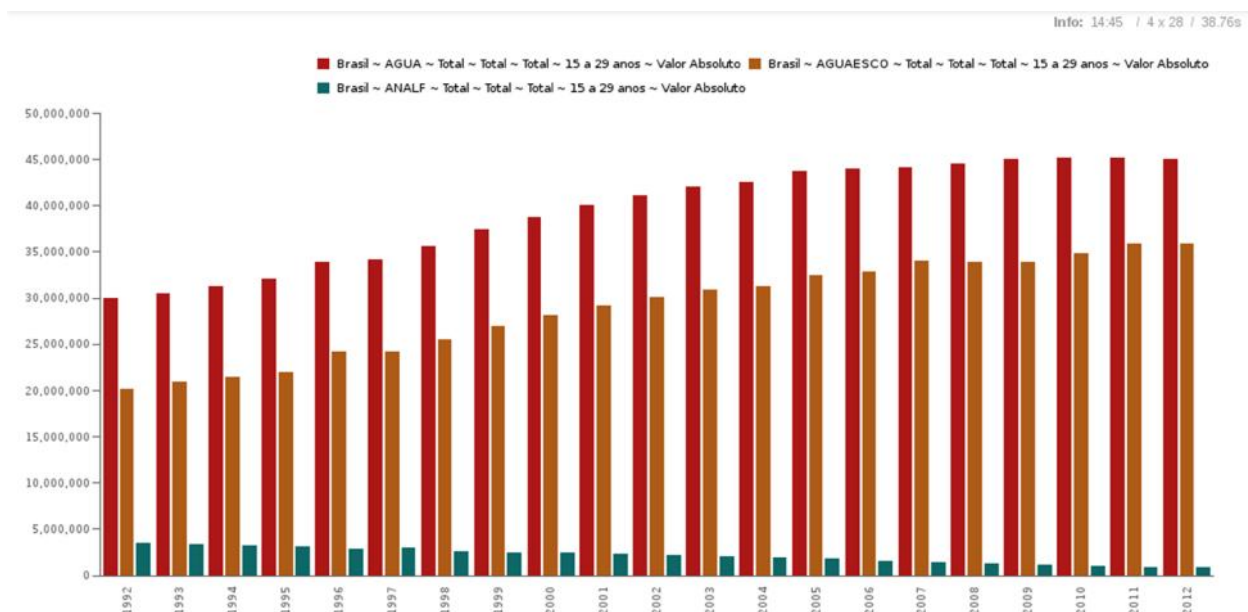
Fonte: Elaborada pelos autores (2017).

Ainda na figura 3, após a extração, tem-se o processo de transformação dos dados. Nesta etapa todas os dados são convertidos para um formato e estrutura comum. Em seguida, tem-se a parte da carga em que realiza-se a inserção de dados em um banco de dados relacional com uma estrutura específica para a criação do cubo.

Posteriormente, o PSW-MONDRIAN foi utilizado para a criação do cubo, o qual permite análise multidimensional de dados podendo-se desenvolver diferentes soluções para exploração do repositório de dados por meio de uma ferramenta de análise para a produção de informações em diversos formatos, tais como, tabela, gráficos e relatórios.

Por fim, utiliza-se o SAIKU para a visualização dos dados conforme a orientação do cubo possibilitando a busca de informação de forma dinâmica. A figura 5 mostra um exemplo em que foram relacionados três bases de dados do CKAN, relacionando jovens com acesso adequado a água, saneamento e sua relação com a taxa de analfabetismo. A informação extraída deste gráfico mostra que jovens com acesso a água e saneamento apresentam taxa de analfabetismo menor.

Figura 5 - Exemplo de relacionamento de conjunto de dados da SNJ.

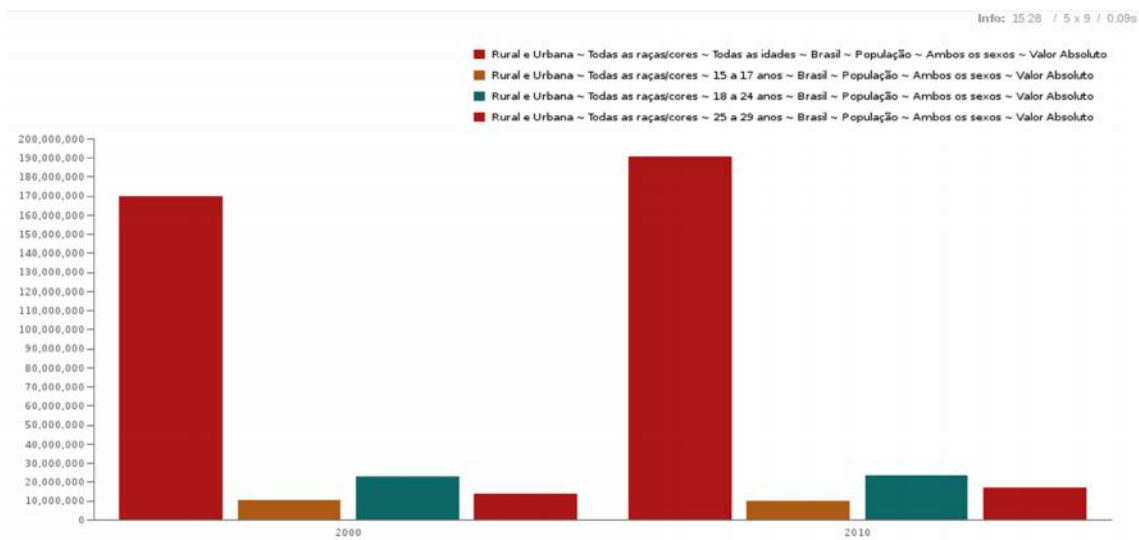


Fonte: Elaborada pelos autores (2017).

De acordo com a figura 5, o gráfico relaciona jovens com acesso a água (barra vermelha), jovens com acesso a saneamento (barra laranja) e sua relação com a taxa de analfabetismo (barra azul) entre os anos de 1992 até 2012. Pode-se observar que quanto mais acesso a água e saneamento menor é a taxa de analfabetismo dos jovens. O relacionamento apresentado envolve faz uso de duas bases de dados distintas (CKAN) no qual são relacionadas entre si no Pentaho trazendo informações que a olho nu não podem ser interpretadas individualmente no CKAN.

Figura 6: Exemplo de conjunto de dados da SNJ. Figura 6: Exemplo de conjunto de dados da SNJ.

Figura 6 - Exemplo de conjunto de dados da SNJ



Fonte: Elaborada pelos autores (2017).

A figura 6 apresenta outro tipo de relacionamento, associando a população brasileira e suas faixas etárias de acordo com os Censos de 2000 e 2010. Observa-se que além da população total ter aumentado (aproximadamente 12%), percebe-se que a população com faixa etária entre 25 a 29 anos também aumentou quando comparado com o censo de 2000. O que demonstra o envelhecimento da população brasileira. Essas informações, embora sejam oriundas de uma mesma base CKAN, apenas podem ser observadas através dos recursos do Pentaho, facilitando assim a análise dos dados.

Vale ressaltar que as consultas podem ser realizadas de forma dinâmica cruzando vários tipos de dados em diversos tipos de gráficos, relatórios e planilhas. Dessa forma, com o SAIKU, pode-se reutilizar os dados do CKAN para geração de novas informações, incrementando as possibilidades de uso das bases de dados mantidas no repositório.

O modelo apresentado no presente estudo encontra semelhanças na experiência de Tygel (2012), na medida em que utiliza o Pentaho para análise de dados abertos governamentais. De Faria Cordeiro et al (2011) utilizaram o Pentaho para integrar uma base de dados para uso de ferramentas semânticas. Mendonça, Cruz e Campos (2014) utilizaram o Pentaho como parte de um modelo para tratamento de dados governamentais. Dos Santos e da Silva (2014) comparam o

sistema I-GOV com o uso do Pentaho para integrar dados governamentais. Portanto, o Pentaho revela-se apropriado ao apoio à análise de dados governamentais de características estatísticas. O resultado do caso de estudo pode ser encontrado em na Plataforma de Dados Abertos de Juventude¹.

4 CONSIDERAÇÕES FINAIS

O objetivo do projeto de pesquisa é apresentar um modelo de uso de ferramentas livre para disseminar dados de governo de forma organizada, a fim de facilitar o seu uso. Da mesma forma, ofertar ferramentas que facilitem o seu uso, como oferta de analisadores de dados. Com isso, o modelo baseado em CKAN e Pentaho atende as necessidades requeridas pela SNJ.

A contribuição do modelo utilizado no presente estudo está no uso de duas ferramentas livres (CKAN e o Pentaho) para atender as necessidades da SNJ para com às orientações dos dados abertos governamentais e ofertar aos gestores de políticas de juventude bem como a sociedade em geral, um sistema web com oferta de dados brutos ou pré-analisados. O modelo desenvolvido e implementado na SNJ composto pelo CKAN e Pentaho apresentou-se eficaz, dinâmico, podendo ser implementado em outros órgãos de governo, contribuindo com a discussão do uso de ferramentas livres em órgãos públicos. Além de se apresentar eficaz no atendimento aos objetivos do estudo. O CKAN possibilita a criação de um repositório, com a catalogação de bases de dados de forma organizada. Para facilitar a recuperação o Pentaho possibilita cruzar dados das bases armazenadas no CKAN para a integração e elaboração de novas informações.

Os recursos tecnológicos do modelo, que apoiam a gestão estratégica, apresentou-se adequados à geração de indicadores que apoiam a tomada de decisões organizacionais, visto que tratam os dados numa integração de gestão e análise. Com isso, tem-se um fluxo composto por depósito, disseminação e análise de dados, expandido os modelo tradicionais baseados somente na gestão de bases de dados.

A metodologia dos dois sistemas combinados é capaz de fortalecer o plano de atuação das organizações por meio da geração de informações rápidas, precisas e personalizáveis garantindo uma estrutura de gestão diferenciada para a melhora no processo de tomadas de decisões pelos

¹<http://magonia.ibict.br/ckan/consultas-livres>.

gestores organizacionais. Por utilizar ferramentas livres tem-se maior possibilidade de continuidade, com sustentabilidade baseada na contribuição da sociedade e do conhecimento dos participantes.

Como trabalhos futuros, pretende-se elaborar um modelo sistemático de integração CKAN-Pentaho para a implementação em demais órgãos do governo contribuindo com a transparência e a disseminação dos dados abertos governamentais a população.

Use of open source software for dissemination and analysis of open government data

ABSTRACT

The study presents a model for the dissemination and analysis of data from government agencies with open source technology. We adopted a methodology focused on the selection of robust technologies that would meet the demands of the agencies involved. The model was based on the Comprehensive Knowledge Archive Network (CKAN) and Pentaho software, which provide data warehousing, retrieval, visualization and analysis services. The case study was applied at Secretária Nacional Juventude, where data from various sources were collected, processed and published for the general public. The model corresponds to the objectives of government agencies, such as the offer of access and analysis of data by the general public, contributing to the transparency of their respective data.

Keywords: Open source. Open data. Data warehouse.

REFERÊNCIAS

AFONSO, Jorge Miguel Dia. **Optimização de estruturas multidimensionais de dados em ambientes OLAP**. 2009. 193 f. Tese (Doutorado) - Curso de Ciências e Tecnologias de Informação, Especialidade em Sistemas Integrados de Apoio à Decisão, Departamento de Ciências e Tecnologias de Informação, Instituto Universitário de Lisboa, Lisboa, 2009.

AHISHAKIYE, Emmanuel et al. Comparative analysis of open source business intelligence tools for crime data analytics. **International Journal Of Latest Research In Engineering And Technology**, Budaun, v. 3, n. 4, p.60-65, abr. 2017.

ALBANO, Cláudio Sonaglio. **Dados governamentais abertos**: proposta de um modelo de produção e utilização de informações sob a ótica conceitual da cadeia de valor. 2014. Tese (Doutorado em Administração) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2014. doi:10.11606/T.12.2014.tde-03062014-170642. Acesso em: 2017-12-08.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regulamento Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. **Diário Oficial da União**. Brasília.

CASTERS, Matt; BOUMAN, Roland; DONGEN, Jos van. **Pentaho Kettle solutions: building open source ETL solutions with Pentaho Data Integration**. [S.l.]: John Wiley & Sons, 2010. 720 p.

CKAN. **Documentation**. 2013. Disponível em: <<http://docs.ckan.org/en/latest/>>. Acesso em: 18 jul. 2017.

CONEGLIAN, Caio Saraiva; SANTAREM SEGUNDO, José Eduardo. Profissional da Informação no contexto dos Dados Abertos: o uso do CKAN para a disponibilização e a organização de dados. **Informação@ Profissões**, v. 5, n. 2, p. 55-78, 2017.

CORDEIRO, Kelli de Faria et al. An approach for managing and semantically enriching the publication of Linked Open Governmental Data. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 24., 2011, Florianópolis. **Anais...** . Florianópolis: Sbc, 2011. p. 82 - 95.

CORRÊA, Andreiuid Sh et al. Transparency and open government data: a wide national assessment of data openness in Brazilian local governments. **Transforming Government: People, Process and Policy**, v. 11, n. 1, p. 58-78, 2017.

COSTA, Evandro B. et al. Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. **Computers in Human Behavior**, v. 73, p. 247-256, 2017.

COSTA, Lucas Rodrigues et al. **Guia do usuário CKAN**. IBICT. 2017.

CURY, Beto. Admirável mundo novo. **Revista Outrolhar**, Belo Horizonte, v. 5, n. 6, p.12-18, nov. 2007. Revista de Debates Mandato Vereador Arnaldo Godoy. Disponível em: <<http://arnaldogodoy.com.br/wp-content/uploads/2014/04/revista-6-juventude-novembro-2007.pdf>>. Acesso em: 12 ago. 2017.

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. **Dados abertos conectados: em busca da web do conhecimento**. [S.l.]: Novatec Editora, 2015.

JARDIM, José Maria; MARCONDES, Carlos Henrique. Políticas de informação governamental: a construção de governo eletrônico na administração federal do Brasil. **DataGramZero: Revista de Ciência da Informação**, v. 4, n. 2, 2003.

KEHL, Maria Rita. A juventude como sintoma da cultura. In: NOVAES, Regina & VANNUCHI, Paulo (orgs.). **Juventude e sociedade**: trabalho, educação, cultura e participação. São Paulo: Perseu Abramo, 2004.

LAMAS, Alejandro et al. Creación de un módulo espacial OLAP para Saiku. In: JORNADAS DE SIG LIBRE, 7., 2014, Girona. **Proceeding**. Girona: SIGTE, 2014. Servei de Sistemes d'Informació Geogràfica i Teledetecció. Disponível em: <http://www.sigte.udg.edu/jornadassiglibre2013/uploads/articulos_13/a26.pdf>. Acesso em: 03 set. 2017.

MARINHEIRO, Antonio; BERNARDINO, Jorge. Experimental evaluation of open source business intelligence suites using OpenBRR. **IEEE Latin America Transactions**, v. 13, n. 3, p. 810-817, 2015.

MENDONÇA, Rogers Reiche de; CRUZ, Sérgio Manuel Serra da; CAMPOS, Maria Luiza M.. Gerência de proveniência multigranular em linked data com a abordagem *ETL4LinkedProv*. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 29., 2014, Curitiba. **Proceeding**. Curitiba: BBD, 2014. p. 87 - 96. Disponível em: <<http://www.inf.ufpr.br/sbbd-sbsc2014/sbbd/proceedings/artigos/pdfs/SBBD2014-Proceedings.pdf>>. Acesso em: 13 set. 2017.

OLIVEIRA, Carolina de. A gestão arquivística de documentos como apoio à publicação de dados governamentais abertos. **Acervo**, v. 29, n. 2, p. 168-178, jul-dez. 2016.

OLIVERIO, Marcio Araujo. Governo aberto como ferramenta de comunicação entre o governo e o cidadão. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 34., 2011, Recife. **Anais...** . Recife: Intercom, 2011.

PALAZZI, Daniele; TYGEL, Alan. Visualização de dados estatísticos representados como dados abertos ligados. 2014. Disponível em: <http://cirandas.net/articles/0027/9811/relatorio_-_Visualiza%C3%A7%C3%A3o_de_Dados_Estat%C3%ADsticos_Representados_como_Dados_Abertos_Ligados.pdf> Acesso em: 10 set. 2017

SALLAM, Rita L. et al. Magic quadrant for business intelligence and analytics platforms. **Gartner RAS core research notes**. Gartner, Stamford, CT, 2014.

SANTOS, João Paulo Clarindo dos; SILVA, Fábio José Coutinho da . IGOV: um sistema de integração de dados governamentais. **Revista Brasileira de Administração Científica**, v. 5, n. 2, p. 8-16, 2014.

SILVA, José Fernando Modesto da . SOFTWARE LIVRE: MODELOS DE SELEÇÃO COMO SUBSÍDIO À GESTÃO BIBLIOTECÁRIA. In: Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação, 22, 2007, Brasília - DF. **Anais...** . Igualdade e Diversidade no Acesso à Informação: da Biblioteca Tradicional à Biblioteca Digital. São Paulo: FEBAB, 2007.

TSUKUMO, Alfredo N. et al. Qualidade de software: visões de produto e processo de software. **II ERI-SBC**, 1997. Disponível em:

<<http://www.felipeximenes.xpg.com.br/arquivos/Qualidade%20de%20Software.pdf> > Acesso em: 20 out. 2017.

TYGEL, Alan. Representação e Visualização de dados estatísticos: os desafios dos dados abertos ligados. 2012. Disponível em:

<http://cirandas.net/articles/0027/9806/os_desafios_dos_dados_abertos_ligados.pdf >. Acesso em: 10 set. 2017