



VII CONGRESSO MUNDIAL DE
ESTILOS DE APRENDIZAGEM

Portal De Dados Estatísticos Do Nordeste Transmontano

Pedro Oliveira
Instituto Politécnico de Bragança
Bragança, Portugal
poliveir@ipb.pt

Paulo Alves
Instituto Politécnico de Bragança
Bragança, Portugal
palves@ipb.pt

Resumo

Este artigo apresenta um projeto que visa resolver a lacuna da centralização de informação estatística referente aos municípios da região do nordeste transmontano de Portugal. Esta informação apesar de existir nas diversas entidades, encontra-se dispersa por diversos locais, alguma dela não se encontrando em formato digital, o que dificulta o processo de consulta e pesquisa para os utilizadores. O conceito de Open Data vai de encontro à massificação do acesso à informação de forma livre e acessível ao utilizador, para que este possa utilizá-la e até mesmo republicá-la sem restrições de direitos de autor. Exemplos de utilização deste conceito, são o projeto Dataviva desenvolvido no Brasil, o European Data Portal com os dados estatísticos da Europa e também o observatório da complexidade económica que possui os dados do comércio internacional. Este projeto foi desenvolvido utilizando a ferramenta open source de catalogação, edição e visualização de dados denominada DKAN.

Palavras-chave: Open-Data, Dataset, Dkan, Estatística, Nordeste

1 Introdução

Este projeto que visa resolver a lacuna da centralização de informação estatística referente aos municípios da região do nordeste transmontano de Portugal, foi desenvolvido utilizando a ferramenta open source de catalogação, edição e visualização de dados denominada DKAN. A mesma tem integração na aplicação de gestão de conteúdos (CMS) Drupal. Um CMS pode definir-se como uma aplicação que permite criar, editar, gerir e publicar conteúdo de forma organizada permitindo que o mesmo seja modificado, removido e adicionado com relativa facilidade. Assim é possível a criação de um portal com as componentes de open data, assim como todos os restantes conteúdos que se pretendam demonstrar, de forma fácil, com um desempenho confiável e elevada segurança disponibilizada pelas constantes atualizações do Drupal.

Os dados na ferramenta DKAN são introduzidos como datasets de informação, possuindo cada dataset tantos recursos quanto os necessários. A informação de cada recurso pode ser introduzida em diversos formatos pelo DKAN, (pdf, xls, csv, json, jpg, png, html, xml, etc.). Consoante a tipologia e formato da informação introduzida, a mesma pode ser visualizada pelo utilizador da plataforma de diferentes formas, entre as quais um gráfico dinâmico configurado em tempo de execução pelo utilizador, consoante a informação que pretenda visualizar nos diferentes eixos. No caso de informação georreferenciada, a mesma pode ser visualizada referenciada no mapa.

Para o processo de introdução de recursos, foi necessário um pré-tratamento da informação. Tendo sido os recursos introduzidos em formato tabular csv, o que permite que o utilizador possa aproveitar a visualização gráfica dos dados, selecionando para cada eixo a variável de informação do recurso que pretenda. Foi ainda dividida a informação em áreas, pois é esta metodologia que a ferramenta DKAN sugere, para tal foram criadas nove áreas (Turismo, Justiça, Saúde, Proteção Social, Demografia, Emprego e Mercado de Trabalho, Educação, Cultura e Desporto, Ciência e Tecnologia e Economia). A informação recolhida foi compilada em duzentos datasets, divididos consoante a área correspondente.

2 Análise do estado de arte

No início do desenvolvimento deste projeto, foi realizada uma análise exaustiva deste tipo de projetos. No sentido de se tomar conhecimento do trabalho que se está a realizar atualmente na área do open data, tanto a nível de práticas e metodologias, assim como ao nível das tecnologias utilizadas para atingir os objetivos do projeto (Gurstein, 2011).

Foram identificados dois casos de estudo essenciais e relevantes para este projeto, tendo estes a seguinte designação: Dataviva e European Data Portal. Foram ainda analisados outros portais Open Data que tiveram um papel importante na análise do estado de arte (Janssen, Charalabidis, & Zuiderwijk, 2012).

O portal DataViva (Hidalgo, 2014) é uma ferramenta de visualização que disponibiliza dados oficiais sobre exportações, atividades económicas e regiões de todo o Brasil, possuindo mais de 100 milhões de tipos de visualizações distintas.

Através destas, o DataViva fornece informações valiosas para a promoção das exportações, decisões de investimento e formulação de políticas públicas. Além disto, ao fornecer fatos sobre o desenvolvimento económico e educação, localidades e suas respectivas oportunidades, o

DataViva ajuda a criar uma base sólida para o diálogo entre o setor público e o privado, fundamentada por dados empíricos (Caro, 2014).

O Portal European Data Portal, trata-se de uma resposta recente por parte da União Europeia, que pretende compilar toda a informação disponível sobre os estados membros. E assim num único local, conseguir fazer-se um estudo sobre os diversos temas, e possibilitar a sua análise por estado membro (Joseph, 2015). Será também possível ter uma perspetiva histórica da evolução das diferentes realidades na Europa. Este portal, foi iniciado há vários meses, e apesar de já ter perto de meio milhão de datasets de informação disponíveis, ainda se encontra em fase experimental, o que demonstra a sua elevada dimensão. A disponibilização da sua versão final está, no entanto, prevista para breve.

Foi também analisado o portal do observatório da complexidade económica que possui os dados do comércio internacional, assim como outros portais de Open Data, nomeadamente os portais das seguintes entidades e regiões: Serra Leoa, Cologne (Alemanha), Porto Rico, Nações Unidas, Louisville (EUA), Arábia Saudita, etc. Estes portais possuem uma dimensão média, e com mais ou menos funcionalidades associadas, contextualizam a informação da região ou entidade para onde foram desenvolvidas.

3 Ferramentas utilizadas

Para o desenvolvimento deste projeto, e após analisado o estado de arte referente ao tema. Foram selecionadas pela sua versatilidade, funcionalidades e atualidade tecnológica as ferramentas CKAN, DKAN e Drupal. As quais são analisadas seguidamente.

3.1 CKAN

A ferramenta CKAN (Winn & others, 2013), foi utilizada como intermédia. Pois foi sentida a necessidade de realizar a introdução massiva de datasets já existentes, tratando-se este de um processo relativamente moroso no caso da sua introdução individual. Assim utilizando a API da plataforma CKAN, e recorrendo à linguagem *Python* foi criado um script que permite a introdução massiva de datasets no espaço de tempo que normalmente seria apenas introduzido um.

A API desta ferramenta permite ainda o denominado processo de *harvesting*. O que possibilitou o processo de migração de informação entre a plataforma CKAN e a plataforma DKAN em poucos minutos.

De salientar que este processo é apenas utilizado quando exista a necessidade de introdução de informação de forma massiva. Estando esta plataforma apenas disponível para a administração do Portal. Para a introdução manual e visualização dos datasets publicamente foi utilizada a plataforma DKAN, que será descrita seguidamente.

3.2 DKAN

A plataforma DKAN (Nuivicic, 2016), derivado do seu processo de atualização constante, das funcionalidades já existentes e das que vêm sendo introduzidas a um ritmo crescente, foi a escolhida para possibilitar a visualização e edição de datasets ao público.

Esta plataforma permite também uma correta integração com a plataforma de gestão de conteúdos (CMS) Drupal.

Ao nível da introdução e visualização dos datasets existentes, é possível a realização de gráficos dinâmicos com a informação disponível, é também possível a disposição da informação em formato tabular. Assim como a visualização desta contextualizada no mapa, quando exista a informação das coordenadas na informação dos datasets. Todo este dinamismo será uma mais valia relevante para consulta da informação por parte do público em geral. Toda a informação disponível poderá assim ser descarregada nos diversos formatos existentes. Na figura seguinte, são ilustradas algumas das funcionalidades de destaque desta plataforma.

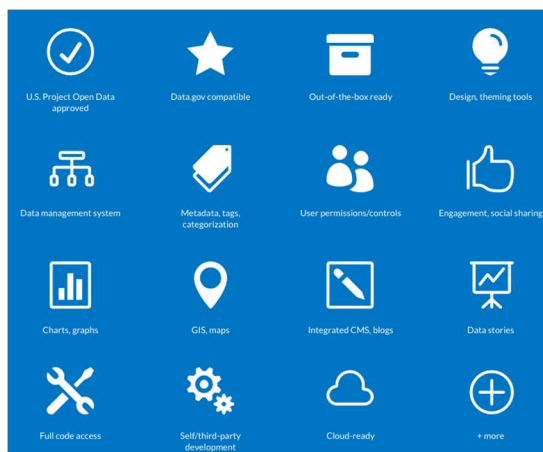


Figura 1 – Funcionalidades DKAN (Nuivicic 2016)

3.3 DRUPAL

Drupal (Drupal, 2016) é um software de gestão de conteúdos (CMS) que se encontra em constante evolução, sendo bastante valorizados pelos utilizadores do sistema, os seus módulos e a flexibilidade, simplicidade e extensibilidade na programação do CMS.

O Drupal contém uma estrutura básica e a interface é simples. As características principais permitem a criação de um site de raiz com o seu conteúdo, grupos de utilizadores e restantes funcionalidades seja feita rapidamente. Para restantes funcionalidades necessárias existe a possibilidade de instalação de add-ons externos que se integram perfeitamente na plataforma, e que permitem a realização das funcionalidades extras que o utilizador necessite. A plataforma Drupal, é assim atualmente um dos mais utilizados gestores de conteúdo mundial. Sendo o seu número de funcionalidades, estabilidade e nível de segurança cada vez mais elevado. Permitindo assim o desenvolvimento base de qualquer tipo de plataforma e a integração com outro tipo de tecnologias, recorrendo a módulos desenvolvidos para integrarem com o Drupal. Este projeto recorreu ao Drupal como plataforma base, onde foi integrado o DKAN. Toda a gestão de utilizadores, menus, conteúdos, e informação auxiliar foi desenvolvida tirando partido da plataforma Drupal.

4 Desenvolvimento do Portal

Para além das funcionalidades de visualização de informação referentes aos datasets já identificadas anteriormente. Foram ainda desenvolvidas três funcionalidades extra que darão suporte a necessidades personalizadas, sentidas no desenvolvimento deste projeto. Foram elas os Dashboards, Cenários e B.I. Estas funcionalidades são descritas em detalhe de seguida.

4.1 Dashboards

Esta funcionalidade ilustrada na figura 2, foi desenvolvida no sentido de possibilitar ao utilizador uma perspetiva rápida dos fatores mais importantes referentes à região. Este tipo de informação está contextualizada localmente e temporalmente, e é atualizada dinamicamente recorrendo à informação inserida na plataforma.

Dashboard

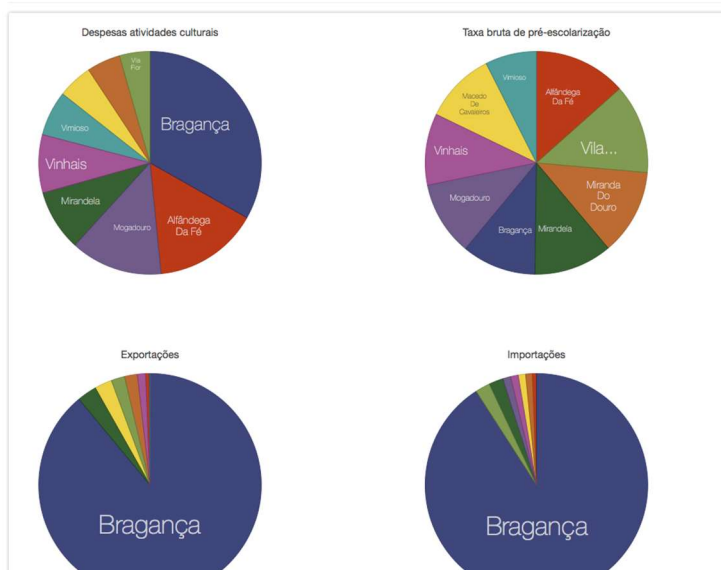


Figura 2 – Dashboards

4.2 Cenários

Esta funcionalidade surgiu com a necessidade de traçar um retrato social nas diversas áreas referenciadas para este projeto, nomeadamente economia, demografia, etc. Como exemplo de cenários existentes, podem ser referidos a evolução demográfica, variação do PIB, etc. A Figura 3, demonstra um cenário desenvolvido, que neste caso apresenta a evolução da demografia. Esta funcionalidade, permite diversos tipos de conteúdo embebido, tais como imagens, gráficos, vídeos ou animações. O que torna a sua visualização mais agradável e interativa para o utilizador.

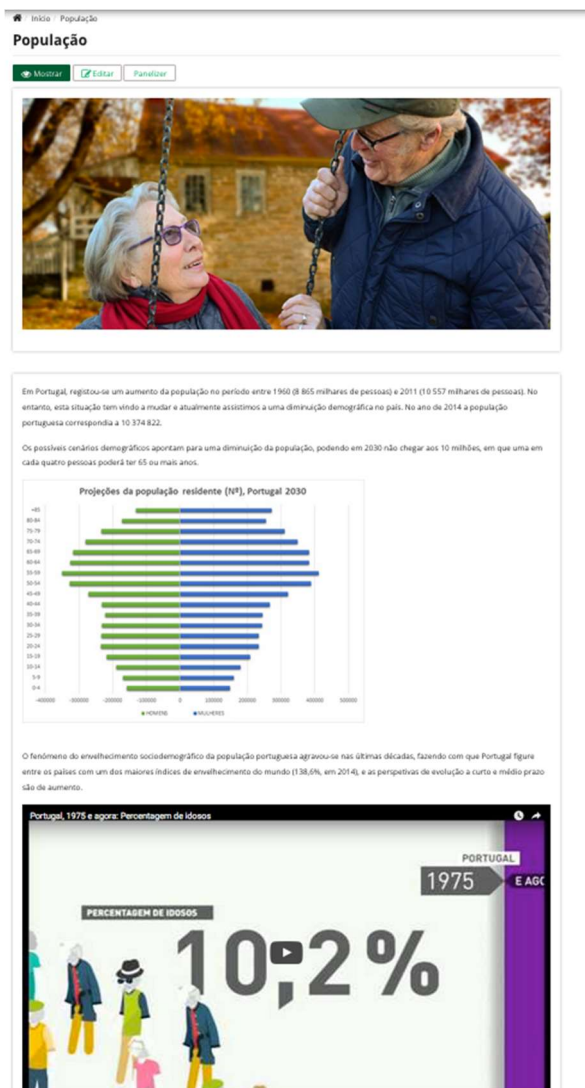


Figura 3 – Cenários

4.3 B.I.

A componente/funcionalidade de maior destaque, será a referente à Business Intelligence. Pois será esta a mais-valia que trará conhecimento extra, resultado do relacionamento da diversa informação disponível. Como pode ser verificado na imagem 4, este tipo de gráficos, estarão disponíveis por área temática, e o utilizador poderá selecionar o tipo de gráfico que pretenda, e também a diversa informação que deseje relacionar. E com isto pretende-se trazer valor acrescido a esta plataforma, e extrair conhecimento da informação, que de forma isolada não seria perceptível à primeira vista.

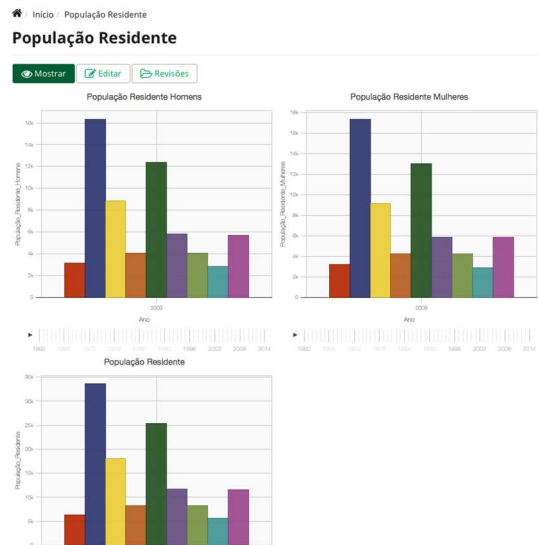


Figura 4 – B.I.

5 Conclusão

Este projeto resultou numa plataforma de informação estatística referente aos municípios da região do nordeste transmontano de Portugal. Como resultados, pode referir-se uma plataforma completamente funcional que para além da visualização dos dados introduzidos estar disponível para toda a comunidade, permite também a inserção de novos datasets, por parte dos utilizadores, consoante as permissões que lhe sejam fornecidas, assim é possível que esta plataforma seja constantemente ampliada, e se mantenha atualizada por parte da comunidade, permitindo assim um fiel retrato económico social da região.

Esta plataforma pode ainda ser vista como uma ferramenta útil de trabalho para investigadores, estudantes, e comunidade em geral uma vez que disponibiliza datasets temáticos que podem ser utilizados em diversos tipos de estudos, incluindo trabalhos científicos.

Este portal, tem a capacidade de armazenar a diversa informação criada em contexto de investigação. E para além disso permitir a sua consulta e visualização no formato adequado ao respetivo tipo de informação. Assim a tarefa do utilizador fica facilitada, e pode a sua atenção focar-se no que realmente interessa, nomeadamente a informação que pretende. Prevê-se que após a divulgação massificada do mesmo, a comunidade académica tenha uma utilização significativa do portal, e tire proveito das suas funcionalidades para maximizar a sua produtividade, o que naturalmente representará melhores resultados na investigação que desenvolvem.

Este tipo de informação contextualizada nas suas diferentes dimensões, pode ser uma fonte de conhecimento de extrema importância para qualquer utilizador, investigador, aluno, ou até mesmo curioso que pretenda ter acesso a este tipo específico de informação, e que ao mesmo tempo será difícil de obter de outra forma. Será ainda solicitado aos diversos investigadores, que realizam investigação referente a temas específicos desta região, que coloquem a sua pesquisa nesta plataforma, assim para além de serem referenciado e reconhecido o seu trabalho, têm a possibilidade de o preservar online de forma segura e gratuita.

O desenvolvimento deste projeto, encontra-se na fase final de conclusão. Estando já disponível ao público como versão de testes. O trabalho futuro, passará agora pela divulgação da plataforma. Serão também realizadas parcerias com as entidades, que poderão disponibilizar a diversa informação da região, consoante a sua área de intervenção. Será ainda perspectivado o desenvolvimento de novas funcionalidades, assim como o aumento do número de datasets e áreas temáticas disponíveis na plataforma.

6 Bibliografia

- Caro, S. (2014). Dataviva. *Government Information Quarterly*, 31, 683.
- Drupal. (01 de 02 de 2016). Drupal. Obtido de Drupal: <https://www.drupal.org/>
- Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16, 2.
- Hidalgo, C. (2014). Big data visualization engines for understanding the development of countries, social networks, culture and cities. *Proceedings of the 25th ACM conference on Hypertext and social media*, 3.
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 258--268.
- Joseph, R. (2015). From data to vision: Big data in government. *Strategic Data-Based Wisdom in the Big Data Era*, 1--14.
- Nuvicic. (01 de 02 de 2016). Nuvicic. Obtido de DKAN: <http://www.nucivic.com/dkan>
- Winn, J., & others. (2013). Open data and the academy: An evaluation of CKAN for research data management.