



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO  
CENTRO DE INFORMÁTICA  
2016.2

# OpenAcademy<sup>UFPE</sup>: Uma ferramenta de Gestão Acadêmica com Dados Abertos Conectados da Universidade Federal de Pernambuco

**Aluno:** Filipe Epifanio de Oliveira (feo@cin.ufpe.br)

**Orientadora:** Bernadette Farias Lóscio (bfl@cin.ufpe.br)

Recife, 13 de Janeiro de 2017

Universidade Federal de Pernambuco

Centro de Informática

Graduação em Sistemas de Informação

# OpenAcademy<sup>UFPE</sup>: Uma ferramenta de Gestão Acadêmica com Dados Abertos Conectados da Universidade Federal de Pernambuco

Trabalho apresentado ao Programa de Graduação em Sistemas de Informação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Recife, 13 de Janeiro de 2017

“ Os olhos do Senhor estão voltados para os Justos, seus ouvidos atentos aos seus clamores.”  
(Sl 34).

## AGRADECIMENTOS

Agradeço a Deus acima de tudo por seu amor e fidelidade, sempre presente e que jamais me deixará. Mesmo diante de minha pequenez perante sua misericórdia ele sempre tem seus olhos em mim, e se cheguei até o grau que cheguei é porque ele antes de todos acreditou em mim.

Agradeço a minha mãe e madrinha Maria mãe de Jesus, que soube dizer sim ao projeto de Deus, sonhou os sonhos de Deus, e se tornou um exemplo de santidade que deseje seguir, obrigado mãe por sua intercessão.

Agradeço a meus familiares que sempre me impulsionaram a ir em frente, a minha mãe Eurides, que mesmo com a morte de meu pai conseguiu se reerguer e lutando criou a mim e meu irmão, mesmo diante de todas as dificuldades que poderíamos enfrentar. A minha família em geral, que mesmo diante de um mundo tão perigoso onde podemos nos perder do caminho correto sempre nos guiou aos caminhos de Deus. Ao meu Pai, que mesmo falecendo tão cedo, no tempo que esteve conosco, sempre nos ensinou a ser um homem de bem. A Clarice, minha namorada por ter aguentado esse tempo de pensamento fixado nos estudos, e por todo apoio que dela recebi nesses dias.

A meus amigos da faculdade, Rayelle que sempre me ajudou muito e sempre gerou admiração pelo seu jeito louco e esforçado de ser uma boa amiga e uma boa estudante. Priscila minha irmã até os 49 do segundo tempo, sou muito grato por ter nessa caminhada. A Jorge, Ruan, Sandrine, Anna Beatriz, Israel, Gerson sempre presente, Carlos um primo mesmo, Hugo, Isadora, João, Bruno, Yasmine, Maurício, Rômulo, Bárbara, Chrystian, Marcelo, Gabriel, Augusto, Cleo que ama mel, Leandro e todos os outros colegas de turma e que passaram em minha história de estudante. A eles serei sempre grato por existirem.

Agradeço ainda e jamais poderia esquecer dos Professores que sempre estiveram presentes e que contribuíram muito para que eu chegasse até aqui, dentre eles José Carlos, José Queiroz, Roberta Madureira, Geber Lisboa, Hermano Perrelli, Kiev Santos, Marcilia Andrade, dentre tantos outros que passaram por minha graduação. Tendo destaque ainda Carla Taciana, que juntamente com minha orientadora recebem uma agradecimento especial, pela presença sempre solícita e disponível a nós alunos e pelo carinho que criei por vocês nesses anos de graduação.

Agradeço a minha orientadora Bernadette Farias Lóscio, por ter acreditado em mim e pela força e carinho com que cuidou de nós seus alunos não apenas no curso do TCC, mas também nas disciplinas cursadas. Sempre levarei no coração “que sou um menino bom”.

Agradecimentos a Rapha, meu chefe que nos momentos que mais precisei ele compreendeu e me liberou para que eu me dedicasse e tivesse um bom

desempenho neste trabalho, e toda equipe do NTI-UFPE, por todo conhecimento que recebi de vocês.

Por fim aos meus irmãos na fé, todos da Comunidade Católica Obra de Maria, aos missionários e aos componentes do Ministério de Música Adoradores de Cristo, que me deram força nos momentos em que me vi preocupado com o meu trabalho..

## RESUMO

A Web, em seu processo constante de evolução, no decorrer dos anos passou a oferecer mecanismos que permitiram aos usuários uma melhor interação com os conteúdos publicados, assim como a possibilidade de contribuir com a publicação de novos dados. Acompanhando o crescimento da Web, a quantidade de dados publicados também vem crescendo em grande escala de volume.

A Web atual destaca-se como um dos principais mecanismos utilizados para o compartilhamento de dados. Entretanto, apesar da facilidade de compartilhamento, em geral, os dados se apresentam de forma desorganizada e não estruturada. Além disso, a maioria das informações podem ser compreendidas apenas por humanos, o que impede o processamento automático por máquinas.

Assim, surge a Web Semântica com o intuito de estruturar e oferecer maior significado aos dados publicados na Web. Para se garantir uma maior organização dos dados, além da estruturação proposta recomenda-se o uso dos princípios de Linked Data, a fim de possibilitar a interligação entre dados disponíveis na Web.

Associado a essa evolução da Web surgiram também conceitos como “Dados abertos” e “Dados Abertos Conectados”, onde, além de estruturados e conectados, deve-se garantir que os dados estejam disponíveis, com suas licenças ao acesso, para que qualquer usuário possa utilizá-los. De maneira geral, a publicação de dados abertos possibilita, dentre outros benefícios, a criação e o surgimento de novas aplicações de consulta e tratamento desses dados.

Levando-se em conta a importância da publicação de dados abertos para organizações públicas, surgiu o interesse pela publicação de dados abertos sobre a Universidade Federal de Pernambuco. Atualmente, encontra-se em desenvolvimento um projeto que tem como objetivo extrair dados do Siga e disponibilizá-los em formato aberto e conectado. Como parte desse projeto, já foi desenvolvida uma ferramenta que permite a publicação automática dos dados do Siga em formato aberto. Além disso, alguns dados relativos a discentes, cursos e disciplinas já estão disponíveis. Neste contexto, este trabalho de graduação tem por objetivo a criação de um portal Web para a realização de análises e geração de relatórios e visualizações sobre discentes, cursos e disciplinas da Universidade Federal de Pernambuco.

## **ABSTRACT**

The Web, in its constant process of evolution, over the years began to offer mechanisms that allowed users to better interact with published content, as well as the possibility of contributing with the publication of new data. Accompanying the growth of the Web, the amount of published data is also growing on a large scale. The current Web stands out as one of the key mechanisms used for data sharing. However, despite the ease of sharing, in general, the data are presented in a disorganized and unstructured way. In addition, most of the information can be understood only by humans, which prevents automatic processing by machines. Thus, the Semantic Web arises in order to structure and offer greater meaning to the data published on the Web. In order to guarantee a greater organization of the data, in addition to the proposed structuring, it is recommended to use the Linked Data principles to provide Interconnection between data available on the Web. In addition to this evolution of the Web, concepts such as "Open Data" and "Connected Open Data" have also emerged, where, in addition to being structured and connected, it must be ensured that data is available, with its access licenses, so that any user can Use them. In general, the publication of open data makes possible, among other benefits, the creation and the appearance of new applications of consultation and treatment of this data.

Taking into account the importance of the publication of open data for public organizations, there was interest in the publication of open data about the Federal University of Pernambuco. Currently, a project is under development that aims to extract data from Siga and make it available in an open and connected format. As part of this project, a tool has been developed that allows the automatic publication of Follow data in an open format. In addition, some data related to students, courses and subjects are already available. In this context, this graduation work aims to create a Web portal for the analysis and generation of reports and visualizations about students, courses and subjects of the Federal University of Pernambuco.

## LISTA DE FIGURAS

Figura 1: Sistemas de 5 estrelas proposto por Tim Berners-lee	16
Figura 2: Ciclo de vida de dados na Web	17
Figura 3: Tela Principal do OpenAcademy <sup>UFPE</sup>	28
Figura 4: Tabela dos Conjuntos de Dados	29
Figura 5: Tela dos Conjuntos de Dados	30
Figura 6: Tela dos Conjuntos de Dados - Downloads	31
Figura 7: Tela de Consultas	32
Figura 8: Resultado da consulta	33
Figura 9: Consulta para Criação do Primeiro Indicador	34
Figura 10: Processamento dos dados para Criação do Primeiro Indicador	35
Figura 11: Indicador de Evasão no curso de Ciência da Computação	36
Figura 12 - Indicador Desistências no Curso de Sistemas de Informação	37
Figura 13 - Padrão de Arquitetura de Software MVC	39
Figura 14: Interação entre OpenAcademy e o DoWMaker	40



## LISTA DE TABELAS

Tabela 1. APIs e Pluguins Geradores de Gráficos	24
Tabela 2. Tecnologias Utilizadas no Desenvolvimento do Portal	41

## Sumário

<b>Capítulo 1. Introdução</b>	<b>11</b>
1.1 Motivação	11
1.2 Objetivos	12
1.3 Estrutura do documento	13
<b>Capítulo 2. Contextualização</b>	<b>14</b>
2.1 Dados Abertos	14
2.2 Dados Abertos Conectados	15
2.3 Ciclo de Vida dos Dados Abertos	17
2.4 Publicação e Consumo dos Dados na Web	19
2.4.1 Publicadores de Dados	19
2.4.2 Consumidores de Dados	20
2.4.3 Formatos de Arquivos para Publicação de Dados	21
2.5 Boas Práticas para a Publicação do Dados Abertos	21
2.6 Visualização dos Dados	23
2.7 Considerações Finais	24
<b>Capítulo 3. O Portal OpenAcademy<sup>UFPE</sup></b>	<b>26</b>
3.1 Motivação	26
3.2 Visão Geral do OpenAcademy <sup>UFPE</sup>	27
3.2.1 Conjunto de Dados	29
3.2.2 Consultas	31
3.2.3 Indicadores	33
3.3 Considerações	37
<b>Capítulo 4. Arquitetura do Portal</b>	<b>39</b>
4.1 Arquitetura do Portal OpenAcademyUFPE	39
4.2 Interação entre o Portal OpenAcademyUFPE e a ferramenta DoWMaker	40
4.3 Tecnologias Utilizadas no Desenvolvimento do Portal	41
4.4 Considerações Finais	42
<b>Capítulo 5. Conclusão</b>	<b>43</b>
Referências Bibliográficas	44

## Capítulo 1. Introdução

Neste capítulo será apresentada a motivação para a realização deste trabalho, assim como os objetivos e a estrutura deste documento.

### 1.1 Motivação

Em seu princípio, a Web em sua versão 1.0, utilizando tecnologias como o HTTP (HyperText Transfer Protocol) e o HTML (HyperText Markup Language) [1], apresentava apenas conteúdos estáticos, o que não permitia uma maior interação entre os usuários e a Web propriamente dita.

Em constante desenvolvimento e apoiado pela popularização da Internet, a Web chegou a sua versão 2.0, que ao quebrar o paradigma da sua versão anterior, foi marcada pelo surgimento de mecanismos que permitiam aos usuários interagir com os conteúdos publicados. Isso contribuiu para um aumento na criação de blogs e sites interativos. Em virtude disso, houve também um crescimento considerável na quantidade de dados publicados e consumidos na Web [2], favorecendo o surgimento de ferramentas de pesquisa e de consulta a esses dados.

Entretanto, em geral, esses dados se encontram de forma desestruturada o que dificulta a realização de consultas mais eficientes e com resultados mais precisos para o usuário, além de uma melhor compreensão dos dados por parte das aplicações. Neste cenário, surge o conceito de Web Semântica. Apresentada como a evolução da Web atual, porém com dados estruturados, e com um significado explícito, possibilitando a compreensão do conteúdo tanto por humanos quanto por máquinas [2].

Associado à Web semântica, foi proposto um conjunto de princípios com o objetivo de não apenas disponibilizar os dados de forma estruturada, mas também de permitir interligação de dados armazenados em diferentes fontes de dados. Então, surge o conceito de “Linked Data” (Dados conectados). Outro conceito que vem ganhando força nos últimos anos é o de “Dados Abertos”. De acordo com [6], dados abertos são dados publicados na Web disponíveis para a utilização de qualquer usuário estando sujeito, no máximo a compartilhar sua licença e creditar a autoria a sua fonte de origem.

Nesse contexto, surge também o conceito de “Dados Abertos Conectados”, possibilitando a reutilização e manipulação de dados conectados e permitindo o surgimento de novas conexões entre os dados existentes [7].

Nos últimos anos, várias empresas e organizações tanto públicas quanto privadas, passaram a investir na publicação de dados abertos. São inúmeros os benefícios da publicação de dados abertos para a sociedade. Do ponto de vista dos dados abertos governamentais, o acesso a esses dados permite à sociedade uma visão mais ampla e transparente da gestão pública, além de incentivar a participação social, geração de emprego, o surgimento de novas aplicações e ferramentas de coleta e tratamento desses dados e o compartilhamento do conhecimento [11].

O Centro de Informática (CIn) da Universidade Federal de Pernambuco (UFPE), vem realizando iniciativas de publicação de dados abertos conectados. Projetos como o *OpenCIn* [2], que é um portal de dados abertos onde são disponibilizadas informações sobre docentes, cursos, entre outras informações sobre o centro e o *OpenSBC* [1], que é um portal que apresenta dados abertos conectados sobre eventos da Sociedade Brasileira de Computação (SBC), recebem destaque entre as iniciativas .

Recentemente, um novo projeto foi iniciado, denominado *OpenUFPE*, em parceria com o NTI, com o intuito de publicar dados abertos da UFPE. Especificamente, o projeto tem como objetivo identificar e extrair dados relevantes do *Sig@*, transformá-los para um formato aberto e então disponibilizá-los de forma pública para a sociedade em geral e, mais especificamente, para a própria comunidade acadêmica da UFPE.

Atualmente, além da ausência de dados disponíveis para a pronta utilização, os gestores acadêmicos da UFPE enfrentam diversas dificuldades para analisar, compreender e gerar indicadores de dados referentes aos discentes, cursos e disciplinas da UFPE. Nesse cenário, este trabalho tem como objetivo o desenvolvimento de um portal que fornece dados abertos referentes aos cursos, discentes e disciplinas ministradas na UFPE. Além da oferta dos dados abertos, o portal permite a realização de análises sobre índices de reprovação, retenção e evasão de alunos, possibilitando geração de relatórios e uma melhor análise sobre os dados para os gestores acadêmicos.

## 1.2 Objetivos

Este trabalho tem por objetivo a criação de um portal Web, denominado *OpenAcademy*<sup>UFPE</sup>, para publicação de dados acadêmicos da UFPE em formato aberto, bem como para auxiliar a gestão acadêmica, permitindo a análise dos dados

sobre indicadores de reprovação, retenção e evasão dos alunos nos cursos e disciplinas ministradas na universidade. Como objetivos específicos do trabalho, destacam-se:

- Publicação de dados abertos acadêmicos da UFPE;
- Especificação de indicadores para auxiliar na gestão acadêmica da UFPE;
- Implementação de um portal para publicação dos dados abertos e dos indicadores para a gestão acadêmica;

### **1.3 Estrutura do documento**

A seguir apresentaremos como este documento está organizado.

- Capítulo 2 : Contextualização
- Capítulo 3 : O portal OpenAcademy<sup>UFPE</sup>
- Capítulo 4 : Conclusão e considerações Finais

## Capítulo 2. Contextualização

Neste capítulo, abordamos os conceitos básicos fundamentais para o entendimento deste trabalho. Na Seção 2.1, apresentamos conceitos referentes a Dados abertos. Na seção 2.2, apresentamos o conceito de Dados abertos conectados. Na Seção 2.3, contemplaremos o ciclo de vida dos dados, Na seção 2.4 apresentamos os processos de publicação e consumo de dados abertos. Nas seções 2.4.1 e 2.4.2 apresentamos a visão dos publicadores e dos consumidores dos dados.

Na seção 2.5 apresentamos as boas práticas para a publicação dos dados abertos e nas seções 2.6 e 2.7 conceitos sobre a visualização dos dados e considerações finais sobre este capítulo.

### 2.1 Dados Abertos

De acordo com a Open Definition (Open Definition, 2014), dados abertos são dados publicados na Web e que podem ser acessados, reutilizados e compartilhados por qualquer indivíduo sem quaisquer restrições de patentes, ou de tipo de usuário, sendo sujeitas apenas à indicação da fonte de origem do dado ou compartilhamento por meio das mesma licenças em que se apresentam a informação [13].

Das características essenciais dos dados abertos podemos citar:

- Disponibilidade e acesso: os dados devem estar disponíveis em um formato acessível de forma que possam ser modificados caso seja conveniente. Os dados devem estar disponíveis de preferência para download através da internet.
- Redistribuição e Reutilização: é necessário que sejam fornecidas condições aos dados de serem reutilizados e redistribuídos permitindo ainda a combinação com outros conjuntos de dados.

- Participação universal: todos sem exceção devem ter a capacidade de utilizar, reutilizar e compartilhar os dados. Não deve haver discriminação contra pessoas, campos de atuação por exemplo.

O especialista em políticas públicas David Eaves [14], elaborou as seguintes leis sobre os dados abertos:

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

Diversas organizações públicas estão investindo na publicação de dados abertos na Web. O Tribunal de Contas da União elaborou uma publicação que especifica 5 motivos para a abertura de dados na administração pública como apresenta [15], apontando as razões para que as organizações públicas realizem iniciativas em abertura de dados governamentais.

1. Transparência na gestão pública;
2. Contribuição da sociedade com serviços inovadores ao cidadão;
3. Aprimoramento da qualidade dos serviços governamentais prestados;
4. Viabilização de novos negócios;
5. Obrigatoriedade por lei.

## **2.2 Dados Abertos Conectados**

A idéia central do conceito de dados conectados (*Linked Data*) tem por objetivo de criar a chamada "Web de Dados." Em outras palavras, refere-se a um conjunto de princípios que são aplicados na publicação e conexão de conjuntos de dados estruturados na Web seguindo os padrões internacionais recomendados pelo W3C[4]. Assim, essas práticas se utilizam de tecnologias presentes na Web, como HTTP (Hypertext Transfer Protocol) e URI (Uniform Resource Identifier), permitindo o processamento automático dos dados por mecanismos de software. Facilitando, dessa forma, a integração dos dados e possibilitando a criação de ferramentas de busca mais eficientes, como navegadores e outros mecanismos. [1]

Os princípios dos Dados Abertos Conectados (Linked Open Data) são:

1. Usar URIs como identificadores para recursos.
2. Usar URIs HTTP para que o usuário possa encontrar esses recursos.
3. Quando alguém acessar uma URI deve-se retornar alguma informação útil no formato de arquivo RDF.
4. Incluir links para outros URIs para que outros recursos possam ser descobertos.

Um fato que se deve levar em consideração é que dados conectados não precisam estar necessariamente abertos, pois uma organização pode prover a conexão de seus dados, mas mantendo-os de forma privada com o objetivo de esconder seus dados. Inclusive há o surgimento um novo conceito chamado “Dados fechados conectados”.

Sobre o termo dados abertos conectados, Tim Berners-Lee propôs ainda uma classificação, conhecida como “Esquema de 5 estrelas dos Dados Abertos”, que classifica por meio de estrelas o grau de abertura em que se encontram os dados.

A Figura 6, representa o sistema proposto de uma forma diferente, elaborada pelo autor:

Figura 1 - Sistema de 5 estrelas proposto por Tim Berners-Lee



Fonte: Elaborado pelo autor

A quantidade de estrelas correspondem:

1. Disponível na internet em qualquer formato possuindo licença aberta. (Por exemplo, no formato PDF)

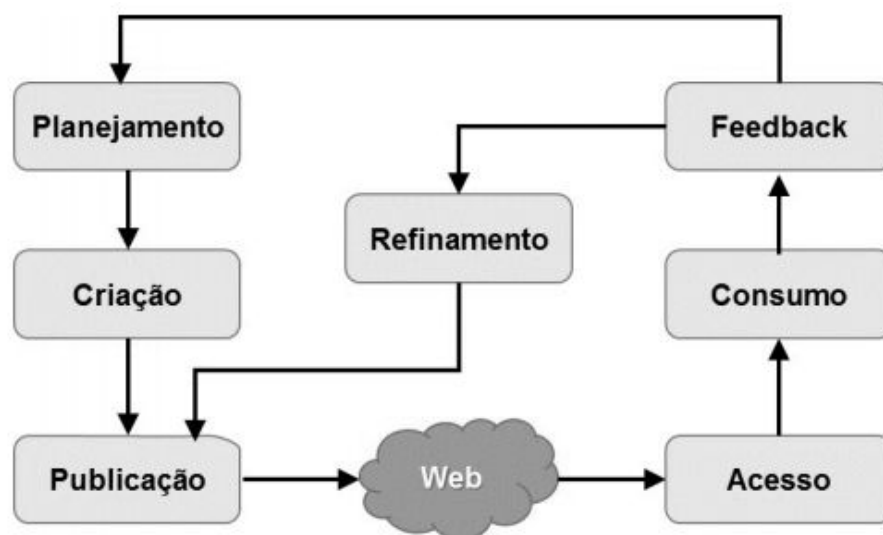


2. Disponível na internet em formato estruturado. Em um arquivo Excel com extensão “XLS”, por exemplo.
3. Disponível na internet de maneira estruturada (CSV ao invés de Excel).
4. Seguindo todas as regras anteriores e os padrões estabelecidos pelo W3C (SPARQL e RDF). Utilizando também URL para identificar os recursos, por onde os consumidores possam direcionar suas consultas.
5. Adiciona todas as regras mencionadas anteriormente, provendo ainda a conexão dos seus dados com outros dados a fim de fornecer um contexto.

### 2.3 Ciclo de vida dos dados abertos

O ciclo de vida dos dados abertos é formado por algumas fases que compõem o processo de publicação e consumo dos dados. Essas fases partem desde a seleção e publicação dos dados até a utilização e o feedback recebido sobre os dados utilizados.

Figura 2 - Ciclo de vida dos dados na Web



Fonte: LÓSCIO, B. F.; OLIVEIRA, M. I. S.; BITTENCOURT, I. I., 2015, p.13.

O modelo proposto é uma instância do Abstract Data Lifecycle Model(ADLM) que foi elaborado por [Möller 2010]. Derivado de uma coleção de modelo de ciclo de vida para domínios centrados em dados, o ADLM é apresentado como um modelo genérico de apresentação de ciclo de vida para dados e metadados [5].

Nesse modelo abstrato um ciclo de vida para domínios centrados em dados possui as seguintes fases: Desenvolvimento de ontologia, Planejamento, Criação, Arquivamento, Refinamento, Publicação, Acesso, Uso externo, Feedback e Término.[5]

O modelo apresentado nesta sessão, não contempla algumas fases do modelo ADLM, como pode ser observado na Figura 2. As fases de Desenvolvimento de Ontologias, Arquivamento e Término não estão contidas no modelo proposto. A fase de Desenvolvimento de ontologias também não foi incluída por se tratar de uma atividade independente. Já as fases de Arquivamento e Término não estão incluídas pois após a publicação na Web os dados já estarão sempre disponíveis.

Observando este modelo, podemos obter uma visão mais compreensiva do processo do ciclo de vida dos dados, desde sua origem até a publicação. A seguir, vamos descrever brevemente as fases propostas no modelo da Figura 2.

- **Planejamento:** Fase que se inicia quando se tem a intenção de publicar os dados até a publicação dos dados propriamente dita. A única observação em relação a priorização dos dados a serem publicados é que se observe o grau de importância que cada dado possui.
- **Criação:** Nesta fase acontece a extração dos dados das fontes correspondentes, ocorre a transformação dos dados nos formatos indicados para a publicação dos dados e são criados os metadados que descrevem os dados.
- **Publicação:** Os dados serão disponibilizados de forma pública na Web. é necessário que os dados estejam atualizados e com toda informação precisa para o bom entendimento dos dados por parte dos consumidores.
- **Acesso:** Os dados já estão disponíveis e podem ser acessados pelos consumidores. Nessa fase, é necessário comunicar que os dados já estão disponíveis e o local onde podem ser acessados.
- **Consumo:** Nesta etapa os dados já estão sendo consumidos. Assim os dados são utilizados para a geração de gráficos e criação de aplicações que realizem análises.
- **Feedback:** A realização do feedback sobre os dados por parte dos consumidores permite identificar pontos de melhorias e possíveis problemas

existentes nos dados. Esta fase é marcada pela coleta dos feedbacks dos consumidores em relação aos dados publicados.

- **Refinamento:** Atividades de manutenção nos dados que foram publicados são realizadas buscando garantir maior segurança aos consumidores.

## 2.4 Publicação e Consumo dos Dados

Com o passar dos anos e a evolução da Web, uma quantidade enorme de dados começou a ser gerada e publicada, e o que antes era apenas uma rede de entretenimento passou a ser o principal canal de compartilhamento de dados, o que motivou diversas organizações a investir na publicação de dados na Web.

A Web, por sua vez, se torna um ambiente onde dois atores principais passam a figurar, os publicadores de dados e os consumidores dos dados, os quais podem ser os próprios publicadores dos dados.

Responsáveis pelo compartilhamento e pela publicação dos dados surgem os publicadores, observando licenças de publicação dos dados, eles se utilizam de mecanismos e ferramentas que possibilitem a manipulação e a visualização dos dados. Os consumidores dos dados, utilizam-se de tais ferramentas para geração de novos dados e de outras informações úteis [1].

### 2.4.1 Publicadores de dados

No cenário de publicação de dados um dos atores que recebe bastante destaque são os publicadores de dados. Eles são responsáveis não apenas pela publicação de qualquer conteúdo, mas por uma publicação de qualidade, legível e compreensível por qualquer consumidor. O publicador de dados, por sua vez, também pode ser o consumidor de seus próprios dados. Outro fator importante que deve ser levado em consideração pelos publicadores são os formatos em que os dados serão publicados, os quais devem ter a capacidade de serem compreendidos não apenas por humanos, mas também por aplicações. Como exemplo de publicadores de dados podemos citar os desenvolvedores de software e o governo.

- **Desenvolvedores de software:** Além de assumirem o papel de consumidores, os desenvolvedores também podem realizar o papel de publicadores de dados. É crescente o número de aplicações que estão sendo

desenvolvidas com o objetivo de coletar dados brutos e realizar a geração de novos dados para assim em seguida publicá-los na Web [1].

- **Governo:** Diversas iniciativas governamentais têm sido levantadas em torno da publicação de dados abertos governamentais. Com o intuito de fornecer uma melhor transparência e respeitando as leis de acesso ao dado, governos de todo o mundo estão investindo na publicação de dados abertos. No Brasil, existem vários portais onde os dados abertos do governo podem ser acessados, entre eles o “Portal de dados abertos do governo federal” do governo brasileiro, que disponibiliza diversos conjuntos de dados governamentais<sup>11</sup>.

#### 2.4.2 Consumidores de dados

Os consumidores também realizam um papel fundamental no contexto de publicação de dados abertos. Em alguns casos, desempenham também o papel de publicadores, para em seguida realizarem o consumo de seus dados visando a geração de outros conteúdos. Podemos citar como exemplo de consumidores de dados os desenvolvedores de software, empresas e organizações e a sociedade em geral.

- **Desenvolvedores de software:** Como mencionamos na seção anterior, os desenvolvedores de software também são responsáveis pela criação de ferramentas que consomem dados da Web, organizam esses dados e por meio dessas aplicações geram novos dados que são disponibilizados para usuários com uma finalidade específica.
- **Sociedade:** Usuários comuns da sociedade que acessam portais de dados abertos e realizam consultas específicas para coletar informações visando uma necessidade específica.
- **Empresas e Organizações:** Esses atores realizam consultas em dados publicados na Web visando obter informações que favoreçam uma tomada de decisão. Podemos citar por exemplo organizações do ramo financeiro, onde informações coletadas a partir desses dados podem representar um grande valor em uma tomada de decisão estratégica da empresa.

### 2.4.3 Formatos de arquivos para a publicação de dados abertos

É um fator fundamental na publicação de dados abertos o fato de que os arquivos estejam disponibilizados em um formato conhecido e estruturado. Segundo [14], é um erro constante em organizações a publicação de dados abertos no formato PDF, pelo fato de ser um formato desestruturado inviabilizando a reutilização dos dados.

Apesar de não ter uma definição formal sobre uma lista de formatos permitidos, alguns formatos de arquivos recebem destaque, pela sua estruturação, por estar num formato de especificação aberta e por ser amplamente conhecido.

A seguir, apresentamos alguns dos formatos mais utilizados na publicação dos dados abertos:

- **Json(JavaScript Object Notation)** - é um formato leve bastante utilizado na troca de dados entre sistemas. Baseado na linguagem JavaScript é um dos formatos mais populares na integração dos dados.
- **XML(Extensible Markup Language)** - é um conjunto de regras baseado em tags muito utilizado para codificar documentos com estrutura hierárquica e em um formato legível por processamento automático. Atualmente vem sendo substituído pelo JSON.
- **RDF(Resource Description Framework)** - modelo de dados estruturado por grafos. Neste modelo os dados são descritos por vocábulos disponíveis na Web. Foi proposto para substituir XML pois o mesmo apresenta limitações de expressividade, além de não permitir que um recurso possa ser descrito de várias formas.
- **CSV(Comma-Separated Values)** - formato para armazenamento de dados de forma tabular, onde cada linha do arquivo representa uma linha na tabela, e as colunas são separadas por vírgula.

## 2.5 Boas práticas para a Publicação de dados na Web

Diante do crescente interesse na publicação dos Dados Abertos, se tornou possível identificar o surgimento de alguns desafios enfrentados por publicadores e consumidores de dados. Neste contexto, o W3C elaborou um grupo de trabalho denominado de Data on the Web Best Practices(DWBP), visando a criação de um guia contendo as melhores práticas para a publicação e consumo de dados na Web.

Assim, foi elaborado um documento intitulado de *Data on the Web: Best Practices, Use Cases & Requirements*. O documento elaborado é um material de apoio que contém algumas boas práticas que auxiliam os provedores de dados a elaborarem conjuntos de dados com qualidade.

De forma resumida, seguem algumas das melhores práticas contidas no documento desenvolvido pelo DWBP:

- **Fornecer metadados:** Os metadados descritivos devem ser disponibilizados com os conjuntos de dados, permitindo um melhor entendimento para os consumidores sobre os dados disponibilizados.
- **Fornecer licenças:** Deve-se especificar ao consumidor se existe alguma restrição para o uso dos datasets disponíveis.
- **Fornecer indicadores de versão:** A fim de assegurar ao consumidor que os dados permanecem atualizados.
- **Utilização de URIs persistentes como identificadores de datasets:** A reutilização de URIs de outros conjuntos de dados aumentam o valor dos dados.
- **Prover dados em Múltiplos formatos:** Pode reduzir custos e a iminência de erros de transformação de dados.
- **Reutilizar vocabulários:** Reduz redundâncias e aumenta a interoperabilidade dos dados.
- **Fornecer os dados para Download:** Favorece a reutilização e o acesso dos dados além de disponibilizar aos consumidores o conjunto de dados completo.

Neste trabalho, fazemos uso de algumas dessas boas práticas. Ao todo, foram elaboradas 35 boas práticas que podem ser acessadas em [12].

A seguir, listamos algumas dessas boas práticas de publicação de dados abertos que foram implementadas No portal OpenAcademy<sup>UFPE</sup> que será detalhado no capítulo 3:

- (Prática 1) Fornecer metadados.
- (Prática 2) Fornecer metadados descritivos.
- (Prática 3) Fornecer metadados estruturais.
- (Prática 7) Fornecer indicador de versão atualizado.
- (Prática 9) URIs persistentes como identificadores de conjuntos de dados.
- (Prática 12) Formatos de dados padronizados e legíveis por máquinas.

- (Prática 14) Fornecer dados em vários formatos.
- (Prática 17) Fornecer download em massa.
- (Prática 20) Acesso em tempo real.
- (Prática 21) Fornecer dados atualizados.
- (Prática 34) Seguir Termos das licenças dos dados.

## 2.6 Visualização dos Dados

Segundo Steele e Iliinsky (2011), a visualização dos dados pode ser vista como um eficiente e eficaz meio de comunicação para um grande volume de informações [8].

Inicialmente, as visualizações eram projetadas de forma manual, o que aumentava o custo do trabalho. Atualmente, as visualizações são projetadas graficamente por meio de codificação ou ferramentas que auxiliem na criação dos diagramas. O uso dessas ferramentas permite uma maior facilidade na atualização dos dados e na inserção de novos dados.

Steele e Iliinsky (2011) apontam ainda alguns motivos que reafirmam a utilidade da visualização dos dados:

- Visualização aproveita as capacidades do sistema visual humano para mover uma enorme quantidade de informações para o cérebro rapidamente;
- Permite identificar padrões, relacionamentos e seus significados;
- Ajuda a identificar subproblemas.
- A visualização é realmente boa para a identificação de tendências, descobrindo pontos específicos ou interessantes em um campo maior.

Nos últimos anos várias APIs foram desenvolvidas visando facilitar a criação de aplicações que promovem a visualização de dados através de uma interface gráfica. Tais APIs fornecem uma gama de facilidades ao desenvolvedor de aplicações Web, dentre elas a possibilidade de gerar gráficos com pouco esforço, assim como a atualização dos dados de forma automática. Dentre essas APIs, podemos citar:

Tabela 1 - APIs e Pluguins geradores de gráficos

<b>Google Charts</b>	é uma API do Google que possibilita a criação de gráficos com baixo esforço. Necessitando apenas de um JavaScript incorporado ao HTML da página da aplicação setando as bibliotecas desta API. Esta API possui vários templates de gráficos
<b>Dojo Charts</b>	API para geração de gráficos estáticos ou interativos no formato HTML5/SVG para aplicações em diversas plataformas como mobile e desktop.
<b>Chart.js</b>	é uma biblioteca JavaScript que Possibilita a criação de gráficos animados no entanto não são interativos. Ideal para realização de relatórios.
<b>Highstock</b>	é uma ferramenta desenvolvida em JavaScript com a finalidade de gerar gráficos em páginas Web possuindo vários templates. Ela se utiliza ainda do jQuery, uma biblioteca Javascript utilizada para simplificar os scripts que se comunicam com o HTML.

Fonte: Elaborado pelo autor

## 2.7 Considerações finais

Apresentamos neste capítulo que consideramos como dados abertos todo dado disponibilizado na Web que são acessíveis sem restrição de usuário, permitindo também a utilização e compartilhamento dos dados por qualquer tipo de usuário. Vimos também que há uma crescente no número de instituições públicas que passaram a investir na publicação de seus dados, isto diante dos benefícios da publicação de dados abertos e diante as leis de liberação de acesso aos dados a todo tipo de usuário.



Para a publicação dos dados abertos, alguns formatos de dados são recomendados diante a estruturação em que os dados serão disponibilizados e seguindo a recomendação das boas práticas de publicação de dados, assim alguns formatos como JSON, XML, RDF, CSV, recebem destaque.

Um outro conceito apresentado foi o de dados abertos conectados que se refere a um conjunto de princípios utilizados na publicação de dados abertos estruturados utilizando tecnologias como HTTP e URI no processamento e busca dos dados.

No contexto de publicação dos dados abertos, observamos que a Web atual compreende como o principal mecanismo de compartilhamento dos dados. Neste cenário estão os publicadores dos dados responsáveis por disponibilizar os dados e os consumidores, que utilizam ferramentas para o processamento e geração de novos dados.

A visualização dos dados abertos através de indicadores permite uma melhor entendimento dos dados publicados, assim várias APIs e ferramentas vêm sendo desenvolvidas para auxiliar a criação de aplicações que gerem indicadores a partir dos dados abertos, neste trabalho citamos algumas delas como o Chart.js e Google chart.

O ciclo de vida dos dados como apresentamos não compreende apenas em selecionar os dados e disponibilizados na Web. Algumas fases compreendem esse processo. Iniciando pelo planejamento, criação dos dados, publicação dos dados, disponibilização de acesso, consumo dos dados, coleta de feedback dos consumidores dos dados e a fase de refinamento dos dados, onde são realizadas melhorias e ajustes nos dados de acordo com o feedback dos consumidores.

Finalizamos este capítulo, apresentando boas práticas utilizadas na publicação dos dados abertos elaboradas pelo W3C, que compreendem a um conjunto de práticas que permitem a publicação dos dados de forma estruturada.

Baseado nas boas práticas de publicação dos dados abertos e nos princípios de dados abertos apresentados nesta sessão, foi criado o OpenAcademy<sup>UFPE</sup>, um portal de apoio ao gerenciamento acadêmico utilizando os dados abertos da UFPE que será detalhado no capítulo 3 deste trabalho.

## Capítulo 3. O Portal OpenAcademy<sup>UFPE</sup>

Neste capítulo, apresentamos o processo de criação do portal de apoio ao gerenciamento acadêmico da UFPE, denominado OpenAcademy<sup>UFPE</sup>. Este portal fornece ao usuário uma série de indicadores educacionais gerados a partir dos conjuntos de dados abertos obtidos do Siga (Sistema Integrado de Gerenciamento Acadêmico) da UFPE.

Na seção 3.1, apresentamos a motivação que nos impulsionou a criação do portal o OpenAcademy<sup>UFPE</sup>. Na sessão 3.2 apresentamos uma visão geral do portal criado. Na seção 3.2.1, listamos os conjuntos de dados contemplados pelo portal. Na seção 3.2.2, detalhamos algumas consultas que podem ser realizadas no portal e os resultados esperados. A seção 3.2.3, apresenta alguns indicadores gerados no portal, que podem auxiliar análises realizadas pelos diretores acadêmicos.

### **3.1 Motivação**

Um coordenador de curso desempenha um papel de grande relevância em uma instituição de ensino. Uma de suas responsabilidades consiste no acompanhamento do desempenho dos alunos nos cursos, observando níveis de reprovação, retenção e evasão dos alunos, permitindo verificar a qualidade dos cursos oferecidos.

Como instrumentos facilitadores para a análise e acompanhamento de tudo o que acontece em um centro acadêmico, os gestores se utilizam de dados e relatórios para a realização de análises e indicadores de desempenho, o que permite uma melhor tomada de decisão diante de eventuais situações.

Na Universidade Federal de Pernambuco, não existe um sistema específico que possibilite ao gestor acadêmico a geração de gráficos e relatórios de forma automatizada. Assim, grande parte dos relatórios são gerados de forma manual.

A partir da necessidade de se ter um sistema específico de apoio ao gerenciamento acadêmico, que dê suporte aos coordenadores de curso e outros interessados na geração de relatórios, indicadores educacionais e realização de pesquisas acadêmicas, foi criado o portal de apoio ao gerenciamento acadêmico, o OpenAcademy<sup>UFPE</sup>.

O OpenAcademy<sup>UFPE</sup> fornece uma interface intuitiva que permite a geração de consultas por parâmetros definidos pelos usuários e a possibilidade de visualizar os dados através de gráficos de indicadores.

### **3.2 Visão geral do OpenAcademy<sup>UFPE</sup>**

O Portal OpenAcademy<sup>UFPE</sup> foi criado com o objetivo de fornecer uma interface mais intuitiva para o acesso e a visualização dos dados abertos da UFPE aos possíveis consumidores de dados, dentre eles os coordenadores de cursos, permitindo ainda a geração de relatórios e a criação de indicadores.

Os dados abertos disponíveis no OpenAcademy<sup>UFPE</sup> são obtidos a partir de APIs de acesso geradas por uma ferramenta elaborada em um outro trabalho de graduação [16], a DowMaker, para a publicação automática de dados abertos do SIG@ na Web. Como estudo de caso, em [12] foram geradas APIs de acesso aos dados relativos à vida acadêmica dos alunos da UFPE. Todo o processo de coleta dos dados para o portal OpenAcademy<sup>UFPE</sup>, é feito via requisição HTTP, através da linguagem de programação JavaScript, linguagem utilizada para a criação do portal e arquivos no formato JSON.

No portal foi criada uma seção de consulta onde o resultado final é a geração de relatórios no formato definido pelo usuário. Nesta seção os consumidores selecionam alguns parâmetros relacionados ao conjunto de dados em que se realizará a consulta, o formato do relatório que será gerado ao término da consulta, além de alguns itens dos conjuntos de dados, como ano do período, data de admissão, entre outros.

Na seção Indicadores, foram criados alguns indicadores de dados educacionais dinamicamente, coletando os dados em tempo real e exibindo na seção indicadores. Através dos indicadores educacionais gerados, os dados poderão ser visualizados graficamente permitindo um melhor entendimento aos consumidores dos dados.

O portal fornece ainda a possibilidade de realizar o download dos conjuntos de dados nos formatos recomendados para a publicação dos dados abertos, JSON, RDF, CSV e XML.

Assim contemplamos os conceitos do “Sistema de 5 estrelas” estabelecido por Tim Berners-Lee, que classifica por meio de estrelas o grau de abertura em que

se encontra os dados, como apresentamos neste trabalho na sessão Dados Abertos.

Apresentamos na Figura 3 a página inicial do portal, onde é descrito de forma sucinta uma breve descrição sobre o portal e em uma tabela, especificamos os datasets contemplados no portal.

Figura 3 - Tela Principal do OpenAcademy<sup>UFPE</sup>



Fonte: OpenAcademy<sup>UFPE</sup>

Na figura 4, apresentamos a tabela que ilustra os conjuntos de dados contemplados pelo portal. Sendo ele Aluno, Curso, Docente, Pessoa, Orgão e Turma. Também foram criados alguns conjuntos de dados específicos, por exemplo AlunoCC, composto por alunos do curso de ciência da computação da UFPE e AlunoSI, formado apenas por alunos do curso de Sistemas de Informação da UFPE. Ao clicar em “Saiba mais”, o usuário será redimensionado para a tela de conjuntos de dados. Nela o usuário poderá verificar informações relacionadas a cada conjunto de dados além de realizar o download do conjunto nos formatos de XML E RDF.

Figura 4 - Tabela dos Conjuntos de Dados

Conjunto de Dados		
#	Nome	Sobre
1	Aluno	<a href="#">Saiba mais</a>
2	Curso	<a href="#">Saiba mais</a>
3	Docente	<a href="#">Saiba mais</a>
4	Pessoa	<a href="#">Saiba mais</a>
5	Orgão	<a href="#">Saiba mais</a>
6	Turma	<a href="#">Saiba mais</a>
7	Aluno SI	<a href="#">Saiba mais</a>
8	Aluno CC	<a href="#">Saiba mais</a>

Fonte: OpenAcademy<sup>UFPE</sup>

Ao selecionar determinado conjunto de dados clicando na aba sobre, o usuário será redimensionado para a tela de conjunto de dados, onde se pode verificar mais informações sobre o conjunto de dados assim como realizar o download dos metadados.

### 3.2.1 Conjuntos de Dados

Seguindo o menu do OpenAcademy<sup>UFPE</sup>, encontramos a sessão conjunto de dados, apresentada na Figura 5. Nela são disponibilizados para download os conjuntos de dados nos formatos recomendados pelas boas práticas de publicação de dados abertos do W3C. Assim, o consumidor pode realizar o download dos dados nos formatos CSV, JSON, RDF e XML.

Figura 5 - Tela dos Conjuntos de Dados

Home

Consultas

Indicadores

Conjunto de Dados

## Metadados

### Descrição do Conjunto de Dados

Título	Pessoa
URI	http://localhost:8080/find/pessoa
Palavras-chave	segurança,automação
Publication date	2017
Criadores	Wilker Santos
Frequência de Atualização	Every Hour
Tema	Segurança
Linguagem	Portuguese
Versão Atual	20170116020019

### Histórico de Versão

Versão	Data de Emissão	URI	Histórico de Mudança
1.0	17-01-2017	http://localhost:8080/find/pessoa/	Criação dos Conjuntos de dados

Fonte: OpenAcademy<sup>UFPE</sup>

A Figura 5, apresenta também o modelo proposto pelo W3C para a tela de publicação e realização de downloads de datasets em aplicações Web, nela apresentamos a descrição de todos os conjuntos de dados presentes no portal. Exemplificamos na figura em questão a descrição do conjunto de dados Pessoa. Nesta seção, ainda descrevemos os formatos dos conjuntos de dados o histórico de versões e por fim, disponibilizamos cada conjunto de dados para downloads nos formatos recomendados para a publicação de dados abertos como apresenta a Figura 6.

Figura 6 - Tela dos Conjuntos de Dados - Downloads

Arquivos para Download				
Título	Descrição	Tipo de Arquivo	Licença	Data de publicação
Aluno.xml	Dataset Aluno em xml	RDF/XML	CC BY-SA 4.0	16-01-2017
Aluno.rdf	Dataset Aluno em rdf	RDF/XML	CC BY-SA 4.0	16-01-2017
Aluno_SI.xml	Dataset Alunos de SI em xml	RDF/XML	CC BY-SA 4.0	16-01-2017
Aluno_SI.rdf>	Dataset Alunos de SI em rdf	RDF/XML	CC BY-SA 4.0	16-01-2017
Aluno_CC.xml	Dataset Alunos de CC em xml	RDF/XML	CC BY-SA 4.0	16-01-2017
Aluno_CC.rdf	Dataset Alunos de CC em rdf	RDF/XML	CC BY-SA 4.0	16-01-2017
Turma.xml	Dataset Turma em xml	RDF/XML	CC BY-SA 4.0	16-01-2017
Turma.rdf	Dataset Turma em rdf	RDF/XML	CC BY-SA 4.0	16-01-2017
Curso.rdf	Dataset Curso em xml	RDF/XML	CC BY-SA 4.0	16-01-2017
Curso.rdf	Dataset Curso em rdf	RDF/XML	CC BY-SA 4.0	16-01-2017
Pessoa.xml	Dataset Pessoa em xml	RDF/XML	CC BY-SA 4.0	16-01-2017
Pessoa.rdf	Dataset Pessoa em rdf	RDF/XML	CC BY-SA 4.0	16-01-2017
Docente.xml	Dataset Docente em xml	RDF/XML	CC BY-SA 4.0	16-01-2017
Docente.rdf	Dataset Docente em rdf	RDF/XML	CC BY-SA 4.0	16-01-2017
Vinculo.xml	Dataset Vinculo em xml	RDF/XML	CC BY-SA 4.0	16-01-2017

Fonte: OpenAcademy<sup>UFPE</sup>

### 3.2.2 Consultas

A Figura 7 apresenta a sessão Consultas, que permite ao consumidor de dados realizar consultas com parâmetros e filtros de igualdade assim selecionados. A interface do portal possibilita ao usuário que ele selecione itens como o dataset em que se deseja realizar a consulta, bem como outros parâmetros relacionados como o nome de um curso, disciplina ou período específico.

Diante dos itens selecionados, a aplicação irá gerar um arquivo no formato JSON que será enviado via requisição HTTP para a API que acessa os datasets já mencionados neste trabalho e retorna o resultado para o OpenAcademy<sup>UFPE</sup>.

A seguir, apresentamos um exemplo de uma consulta realizada no conjunto de dados alunos. É uma consulta simples os alunos matriculados no curso seleciona todos os alunos que estão matriculados no curso de Sistemas de Informação da UFPE, e que foram admitidos no ano de 2016.

Figura 7 - Tela de consultas

Realize sua consulta selecionando os itens dos datasets Limpar Buscar

Datasets	Formatos
Alunos	XML

Select	Valor
Nome do Curso	SISTEMAS DE INFORMAÇÃO
Ano de admissão	2016
Tipo de Situação Acadêmica	Matriculado

Download

Fonte: OpenAcademy<sup>UFPE</sup>

Ao clicar no botão Download, é realizado o download do resultado da consulta, sendo este em formato XML. No entanto o resultado também pode ser obtido no formato RDF. O resultado da consulta em questão pode ser observado na figura 8.



Figura 8 - Resultado da Consulta

```
This XML file does not appear to have any style information associated with it. The document tree is shown below.

▼<XML>
  ▶<aluno_SI>...</aluno_SI>
  ▶<aluno_SI>...</aluno_SI>
  ▶<aluno_SI>...</aluno_SI>
  ▶<aluno_SI>...</aluno_SI>
  ▼<aluno_SI>
    <_id>258797_20170112030031</_id>
    <CNPJ_insttc_ensn>24134488000108</CNPJ_insttc_ensn>
    <codg_area_concmt>null</codg_area_concmt>
    <data_incs_avalc_disct>2016-04-20 00:00:00.0</data_incs_avalc_disct>
    <data_colc_grau_sa>null</data_colc_grau_sa>
    <fim_perd_letv>2016-07-25 00:00:00.0</fim_perd_letv>
    <codg_progm_fomc>311</codg_progm_fomc>
    <nota_enem>660</nota_enem>
    <ano_concs>null</ano_concs>
    <cdg_avalc_disct_turma>8526498</cdg_avalc_disct_turma>
    <codg_tipo_ingss>15</codg_tipo_ingss>
    <codg_tipo_sitc_acadm<1</codg_tipo_sitc_acadm>
    <nota_avalc_disct_turma>9.8</nota_avalc_disct_turma>
    <data_defnc_sitc>2016-08-01 15:19:43.0</data_defnc_sitc>
    <data_fim_real_perd_letv>2016-07-11 00:00:00.0</data_fim_real_perd_letv>
    <codg_turno>5</codg_turno>
    <codg_modl<2</codg_modl>
    <ensn_medio_origm>null</ensn_medio_origm>
    <matcl_curso_v>200873577</matcl_curso_v>
    <matcl_curso_ad<200873577</matcl_curso_ad>
    <obsvc_sitc_acadm<null</obsvc_sitc_acadm>
    <id_compnt_avalc>EE_1</id_compnt_avalc>
    <version>20170112030031</version>
```

Fonte: OpenAcademy<sup>UFPE</sup>

### 3.2.3 Indicadores

Indicadores se apresentam como uma forma intuitiva e mais simples para a representação de dados em geral, pois permitem um entendimento mais eficaz sobre determinado domínio. Assim, os indicadores funcionam como um instrumento de medição que permite diagnosticar situações ou evidências em consequência de determinadas ações já ocorridas. As informações apresentadas pelos indicadores podem auxiliar, por exemplo, em uma tomada de decisão estratégica.

O OpenAcademy<sup>UFPE</sup> fornece na sessão Indicadores uma lista de indicadores pré-definidos com o intuito de auxiliar os coordenadores de cursos no processo de análise de desempenho dos cursos, assim como apresentar aos consumidores comuns um conjunto de dados em uma interface gráfica amigável e de fácil entendimento.

A fim de identificar quais indicadores teriam relevância para os possíveis consumidores dos dados, realizei algumas entrevistas com esses possíveis consumidores. Nesse sentido, entrevistei o responsável pelo núcleo de

acessibilidade da UFPE, e com o Coordenador de Ensino do Cin Geber Ramalho, o qual me sugeriu a criação de alguns indicadores educacionais que servem de apoio a gestão acadêmica. dentre eles podemos citar:

- Disciplinas com maior índice de reprovações na UFPE.
- Média de reprovações por curso.
- Número de evasão de alunos por período em seus respectivos cursos.
- Número de evasão por Curso.

Entretanto, devido a alguns problemas técnicos relacionados a curva de aprendizagem das tecnologias utilizadas na criação do portal e ao tempo utilizado para realizar a integração com a API DoWalker, responsável por disponibilizar os dados, nem todos esses indicadores foram implementados. Dentre os implementados podemos destacar os indicadores que representam o número de evasão de alunos da UFPE.

O indicador a seguir apresenta o número de alunos que desistiram do curso de Ciência da Computação no período de 2012 a 2016. A Figura 9, representa a consulta no formato JSON enviada a API para selecionar os dados que serão utilizados na construção do indicador:

Figura 9 - Consulta para a geração do primeiro indicador

```
$http({
  url: 'http://localhost:8080/find/execute_query',
  method: 'POST',
  data: {

    "datasetName": "aluno_CC",
    "format": "json",
    "where":
    {
      "codg_progm_fomc" : "45",
      "codg_tipo_sitc_acadmc" : "17"
    }
  },
  headers: {
    'Content-Type': 'application/json'
  }
}).then(function successCallback(cc) {
  $scope.indicadores = cc.data;
});
```

Fonte: OpenAcademy<sup>UFPE</sup>

Ao enviar esta requisição a API, recebemos em nosso portal os dados que setados em variáveis onde realizamos verificações do ano de admissão do aluno e a situação acadêmica de cada aluno, como ilustra na Figura 10, se o campo `codg_tip_sitc_acadmc` estiver com o valor 17, representa que o aluno desistiu do curso. E em seguida contabilizamos a quantidade de alunos desistentes e adicionamos o valor em variáveis que serão adicionadas nos campos dos indicadores.

Figura 10 - Processamento dos dados para criação do primeiro indicador

```
for (var i = 0; i < dadosCC.length; i++) {
  if (dadosCC[i].ano_adms == "2012" && dadosCC[i].codg_tipo_sitc_acadmc == "17"){
    desistiuCC2012 = desistiuCC2012 + 1;
    console.log(desistiuCC2012 + "desistentes CC em 2012");
  }

  else if (dadosCC[i].ano_adms == "2013" && dadosCC[i].codg_tipo_sitc_acadmc == "17"){
    desistiuCC2013 = desistiuCC2013 + 1;
    console.log(desistiuCC2013 + "desistentes CC em 2013");
  }

  else if (dadosCC[i].ano_adms == "2014" && dadosCC[i].codg_tipo_sitc_acadmc == "17"){
    desistiuCC2014 = desistiuCC2014 + 1;
    console.log(desistiuCC2014 + "desistentes CC em 2014");
  } else if ( dadosCC[i].ano_adms == "2015" && dadosCC[i].codg_tipo_sitc_acadmc == "17"){

    desistiuCC2015 = desistiuCC2015 +1;
    console.log(desistiuCC2015 + "desistentes CC em 2015");
  }

  else if ( dadosCC[i].ano_adms == "2016" && dadosCC[i].codg_tipo_sitc_acadmc == "17"){

    desistiuCC2016 = desistiuCC2016 +1;
    console.log(desistiuCC2016 + "desistentes CC em 2016");
  }
}

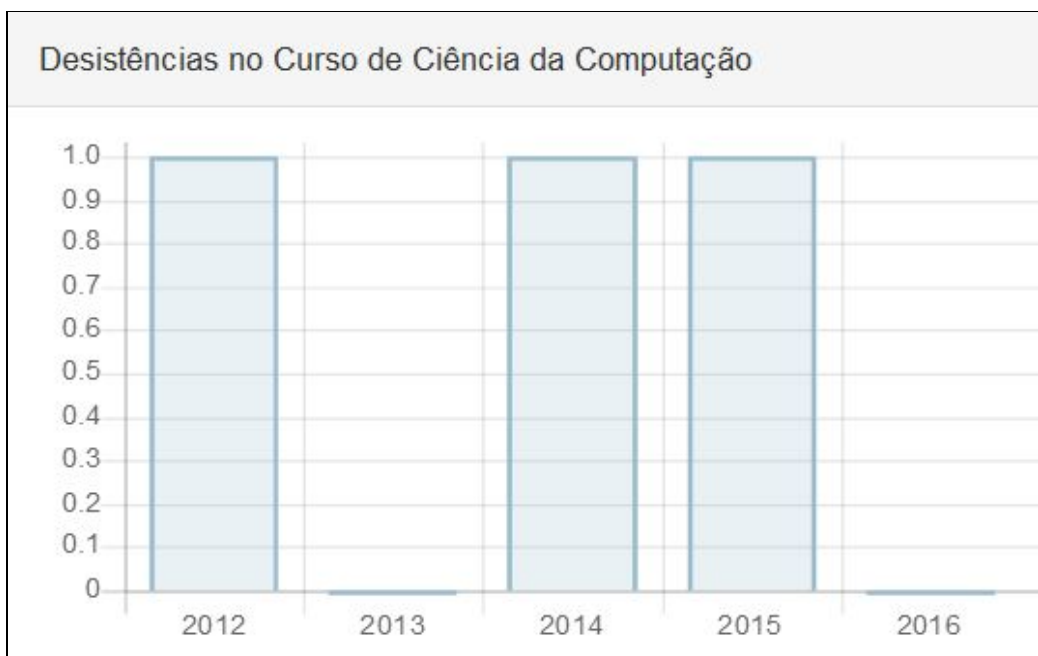
$scope.bar_cc = {
  labels : ['2012', '2013', '2014', '2015', '2016'],
  series : ['Numero de repovacoes no ultimo semestre', 'numero de reprovacoes da disciplina']
  data : [
    [desistiuCC2012, desistiuCC2013, desistiuCC2014, desistiuCC2015, desistiuCC2016]
  ]
};
```

Fonte: OpenAcademy<sup>UFPE</sup>

A Figura 11 apresenta o resultado gerado pela consulta e o processamento dos dados demonstrado anteriormente.

O indicador “Desistências no Curso de Ciência da Computação”, apresenta que dos alunos que ingressaram de ciência da computação no ano de 2012, apenas 1 desistiu do curso, Assim também ocorreu nos anos 2014 e 2015, onde apenas 1 aluno ingressante em 2014 e 1 aluno admitido em 2015, desistiu do curso.

Figura 11 - Indicador Desistências no Curso de Ciência da Computação



Fonte: OpenAcademy<sup>UFPE</sup>

A figura 12 apresenta outro indicador gerado a partir dos dados do SIG@. O indicador “Desistências do Curso de Sistemas de Informação”, representa a quantidade de alunos que abandonaram o curso de Sistemas de Informação da UFPE no período de 2011 a 2016.

Figura 12 - Indicador Desistências no Curso de Sistemas de Informação



Fonte: OpenAcademy<sup>UFPE</sup>

O resultado como apresenta a descrição do indicador, mostra que dos alunos que ingressaram no curso de Sistemas de Informação em 2011, apenas 3 alunos desistiram do curso. Dos alunos ingressantes em 2013, apenas 2 deles abandonaram o curso e dos alunos admitidos em 2014, apenas 1 desistiu do curso.

Tais indicadores permitem aos diretores acadêmicos a possibilidade de realizar uma análise referente às desistências dos alunos nos seus respectivos cursos seguindo a linha do tempo dos anos de admissão de cada aluno.

### 3.3 Considerações

Neste capítulo apresentamos a criação do portal OpenAcademy<sup>UFPE</sup>, portal que servirá de apoio ao gerenciamento acadêmico para os cursos da UFPE. O portal que acessa APIs para realizar coleta de dados acadêmicos do sistema SIG@ e em seguida disponibiliza os dados nos formatos recomendados para a publicação

dos dados abertos e também fornece alguns indicadores educacionais, permitindo aos diretores de curso a realização de análises e obter um entendimento mais eficaz sobre a situação dos alunos e dos cursos da UFPE.

Assim, Ao disponibilizar os metadados dos conjuntos de dados contemplados neste portal, surge também a possibilidade de geração de pesquisas e outras atividades acadêmicas com os dados abertos da UFPE.

## Capítulo 4. Arquitetura do Portal

Neste capítulo, apresentaremos especificações do Portal OpenAcademy<sup>UFPE</sup> em relação a sua implementação. Nele apresentamos alguns detalhes sobre a arquitetura utilizada, as tecnologias e ferramentas utilizadas na implementação e por fim, a interação existente entre o Portal OpenAcademy<sup>UFPE</sup>, a aplicação DoWMaker e os dados do SIG@.

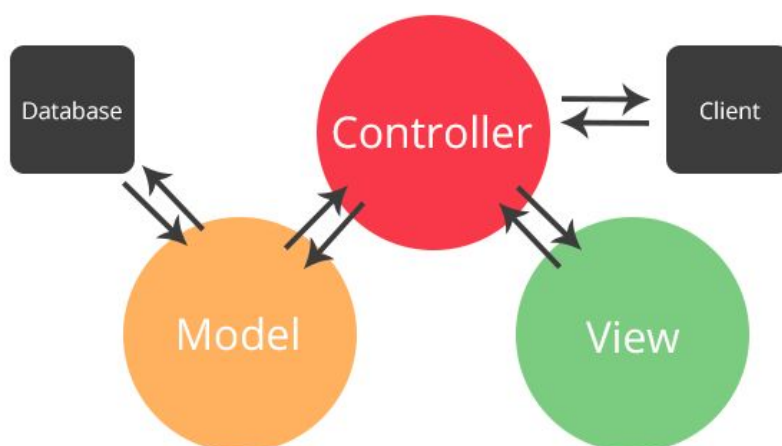
### 4.1 Arquitetura do PortalOpenAcademy<sup>UFPE</sup>

O padrão de arquitetura de software utilizado no desenvolvimento do portal foi o padrão MVC (Model–view–controller). Este padrão de arquitetura de software separa a aplicação em três camadas: view, model e controller.

A camada View é utilizada para realizar a interação da aplicação com o usuário. No portal criado as páginas em HTML são responsáveis pela exibição dos dados para o usuário. A camada Model é responsável pela manipulação dos dados. Responsável sobretudo pela leitura e escrita dos dados e também por realizar validações nos dados existentes. Já a camada Controller, é responsável por receber todas as requisições realizadas pelo usuário. É a camada responsável por direcionar em qual View os dados serão apresentados ao usuário.

A figura a seguir, representa o padrão de arquitetura de software MVC:

Figura 13: Padrão de Arquitetura de Software MVC



Fonte: ASP.NET fundamentals MVC

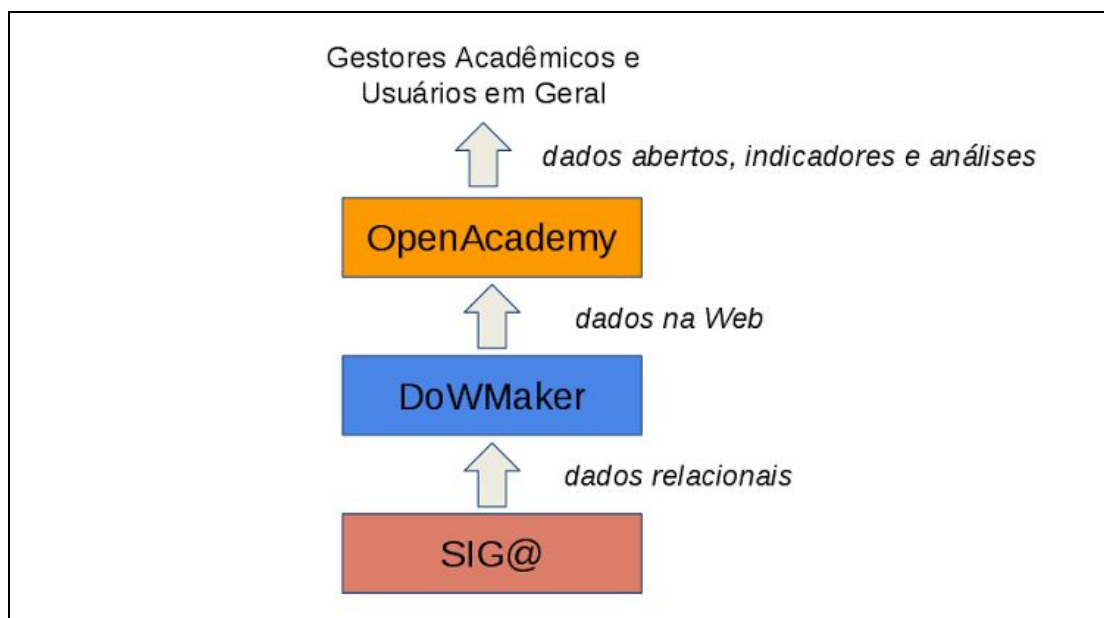
Este padrão de arquitetura de software vem sendo muito utilizado em aplicações Web, sobretudo em aplicações desenvolvidas na linguagem de programação JavaScript juntamente com o framework AngularJS desenvolvido pelo google. Tais ferramentas foram utilizadas na criação do portal OpenAcademy<sup>UFPE</sup> como apresentaremos em breve.

#### 4.2 Interação entre o Portal OpenAcademy<sup>UFPE</sup> e a ferramenta DoWMaker

O portal OpenAcademy interage diretamente com a aplicação DoWMaker. Entretanto, o portal pode ser configurado para acessar qualquer Api que dê permissões para o portal realizar requisições HTTP.

Como apresentamos no capítulo 3 deste trabalho, as requisições realizadas pelo portal utilizam-se do método POST, enviando parâmetros no formato JSON para a API DoWMaker que receberá a requisição, processará os dados e retornará o resultado no formato solicitado na requisição. A Figura 13, representa como se dá a interação entre as aplicações inseridas no contexto deste trabalho:

Figura 14: Interação entre OpenAcademy e o DoWMaker



**Fonte:** DoWMaker: Uma Ferramenta para Publicação Automática de Dados na Web







Dessa forma, os dados são coletados do banco de dados do SIG@ pela API DoWMaker, processados e disponibilizados nos formatos de arquivos solicitados via requisição HTTP, tais requisições são realizadas pelo portal OpenAcademy<sup>UFPE</sup>.



### 4.3 Tecnologias Utilizadas no Desenvolvimento do Portal

Na construção do portal utilizamos algumas tecnologias frequentemente usadas no desenvolvimento de aplicações Web. A tabela 2 apresenta uma descrição das tecnologias utilizadas do desenvolvimento do OpenAcademy<sup>UFPE</sup> :

Tabela 2 - Tecnologias Utilizadas no Desenvolvimento do Portal

	O HTML é uma linguagem de marcação utilizada no desenvolvimento de páginas Web.
	JavaScript é uma linguagem de programação baseada em <i>scripts</i> . Inicialmente criada para como parte dos navegadores Web para que os <i>Scripts</i> pudessem ser executado no lado cliente. Os scripts desenvolvidos em JavaScript são amplamente integrados em páginas Web devido a facilidade de interação com o Document Object Model(DOM) da página.
	AngularJs é um framework JavaScript open-source mantido pela Google, que permite a criação de aplicativos que podem ser acessados por um navegador Web. O AngularJs segue o padrão MVC da engenharia de software e se utiliza o conceito de injeção de dependência fornecendo serviços para o lado do servidor.
	Bootstrap é um framework front-end que possibilita aos desenvolvedores a criação de sites e aplicações com tecnologia mobile (responsivo). Além disso, o bootstrap fornece uma gama de pluquins em JavaScript (jQuery) facilitando ainda mais o trabalho de designers e desenvolvedores.
	É uma linguagem de folhas de estilo, utilizada para definir a apresentação de documentos escritos em linguagem de marcação como HTML. Através do CSS podemos customizar e modificar a aparência das páginas HTML em nosso projeto.
	O Bower é uma ferramenta que auxilia a gestão de dependências de código do front-end. Através de um arquivo denominado “bower.json”, definimos as dependências que serão instaladas no nosso projeto.

Fonte: Elaborado pelo Autor

#### 4.4 Considerações Finais

Neste capítulo apresentamos a arquitetura de software utilizada no desenvolvimento do portal OpenAcademy<sup>UFPE</sup>. Para o portal utilizamos o padrão MVC (Model–view–controller) , padrão que é utilizado pelo AngularJs, uma das tecnologias utilizadas no portal. Apresentamos também a interação existente entre o portal e a ferramenta DoWMaker, desenvolvida em um trabalho de graduação em paralelo a este trabalho. A ferramenta DoWMaker cuida da extração dos dados do banco de dados do SIG@ e na transformação dos dados para a geração de APIs que serão acessadas pelo portal OpenAcademy<sup>UFPE</sup>.

Finalizamos apresentando ainda algumas tecnologias utilizadas no portal. Selecionamos as tecnologias mais utilizadas atualmente por desenvolvedores de aplicações Web, dentre elas podemos citar: Bootstrap, CSS, HTML, JavaScript, Bower e o AngularJs.

## Capítulo 5. Conclusão

A evolução da Web atual vem impulsionando a publicação de dados abertos. Diante da importância da publicação dos dados abertos, organizações governamentais e não governamentais passaram a investir na publicação dos dados tornando este um assunto bastante discutido e bem visto.

Plugins e APIs são utilizados como ferramentas de apoio aos desenvolvedores para realizarem a apresentação dos dados de uma forma mais intuitiva e de fácil entendimento aos consumidores de dados.

Também abordamos conceitos referentes à publicação de dados abertos, assim como boas práticas de publicação dos dados na Web criadas pelo W3C. Mencionamos ainda conceitos sobre o ciclo de vida dos dados e os principais atores presentes no processo de publicação e consumo dos dados.

Como resultado deste trabalho, foi criado o OpenAcademy<sup>UFPE</sup>, um portal de apoio ao gerenciamento acadêmico para a UFPE. Este Portal acessa dados de 5 datasets criados a partir do banco de dados do SIGA UFPE via uma API, e permite a geração de relatórios de consultas e a criação de indicadores.

Dentre as principais contribuições contidas neste trabalho podemos destacar a criação de indicadores educacionais que poderão ser utilizados para realização de análises sobre os cursos, disciplinas e alunos existentes na UFPE.

Como trabalhos futuros, destacamos melhorias no portal, como a implantação de novas técnicas de consumo e visualização dos dados. Dentre as melhorias podemos citar:

- Permitir aos consumidores a possibilidade de salvar os relatórios gerados, bem como visualizar relatórios anteriores.
- Melhorias na interface gráfica, além de a criação de novos módulos visando o consumo dos dados coletados pela API.
- Criar um gerador de Portais.

## REFERÊNCIAS

1. CRUZ, R.INGRID.,V. “**OpenSBC: Um Portal de Dados Abertos Conectados sobre os Eventos da Sociedade Brasileira de Computação**”. Universidade Federal de Pernambuco. Recife, 2016.
2. SILVA, K.M. “**Publicação e Análise de Dados Abertos Conectados sobre os Cursos de Graduação do CIn/UFPE**”. Universidade Federal de Pernambuco. Recife, 2015.
3. LÓSCIO, B.F., FILHO, F.W. “**Web Semântica: Conceitos e Tecnologias**”(2010)
4. ROCHA, J. D. L. “**OpenCIn - Dados Abertos Interligados Acerca dos Docentes do Centro de Informática**”. Universidade Federal de Pernambuco. Recife, 2015.
5. SIMPÓSIO BRASILEIRO DE BANCO DE DADOS 2015 - **TÓPICOS EM GERENCIAMENTO DE DADOS E INFORMAÇÕES 2015**. Disponível em: <<http://dexl.Incc.br/sbbd2015/anais/ShortCourses.pdf>>. Acesso em Outubro: 2016.
6. OPEN KNOWLEDGE INTERNATIONAL. **The Open Definition. Open Definition**. Disponível em: <<http://opendefinition.org/>>. Acesso em: 25 Março 2016.
7. ISOTANI, S.; BITTENCOURT, I. I. **Dados Abertos Conectados**. São Paulo: Novatec, v. 1, 2015.
8. AZEVEDO, P.C.N. “**Visualização de dados abertos vinculados em sistemas de informações geográficas uma revisão sistemática da literatura**”. Universidade Federal de Santa Catarina. LOD BRASIL, LINKED OPEN DATA. Novembro 2015. Disponível em <[https://zenodo.org/record/12549/files/Visualizacao\\_de\\_dados\\_abertos\\_vinculados\\_em\\_sistemas\\_de\\_informacoes\\_geograficas\\_uma\\_revisao\\_sistematica\\_da\\_literatura.pdf](https://zenodo.org/record/12549/files/Visualizacao_de_dados_abertos_vinculados_em_sistemas_de_informacoes_geograficas_uma_revisao_sistematica_da_literatura.pdf)>
9. FILHO, F.W; LÓSCIO, B. F (2010) **Web Semântica: Conceitos e Tecnologias**. Minicurso apresentado na Escola Regional de Computação Ceará - Maranhão – Piauí 2010.
- 10 -SILVA,G.C. “**RDF e RDFS na Infra-estrutura de Suporte à Web Semântica**”. Departamento de Ciência da Computação – Instituto de Ciências Exatas Universidade Federal de Juiz de Fora.
- 11 - **Manual dos dados abertos do governo**. Disponível em: <[http://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual\\_Dados\\_Abertos\\_WEB.pdf](http://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf)>

12 - WORLD WIDE WEB CONSORTIUM. **Data on the Web Best Practices. Data on the Web Best Practices, 2016.** Disponível em: <<https://www.w3.org/TR/dwbp/>>. Acesso em: 19 Outubro 2016.

13 - OPEN DATA HANDBOOK. Disponível em:  
<<http://opendatahandbook.org/guide/en/what-is-open-data/>>

14 - CARTILHA DE PUBLICAÇÃO DE DADOS ABERTOS. Disponível em:  
<<http://dados.gov.br/paginas/cartilha-publicacao-dados-abertos>>

15 - **Cinco Motivos para a Publicação de Dados Abertos na Administração Pública - TCU.** Disponível em: <<http://portal3.tcu.gov.br/portal/pls/portal/docs/2689107.PDF>>

16 - CAVALCANTI, W.R.S. "**DoWMaker: Uma Ferramenta para Publicação Automática de Dados na Web**". Universidade Federal de Pernambuco. Recife, 2017.