



**CENTRO UNIVERSITÁRIO DE BRASÍLIA - CEUB**  
**PROJETO DE INICIALIZAÇÃO CIENTÍFICA**

**PIETRO VICTOR MACIEL**

**UM ESTUDO SOBRE DADOS GOVERNAMENTAIS ABERTOS (DGA) NO  
CONTEXTO DO BRASIL COM O SUPORTE DE CIÊNCIA DE DADOS**

**BRASÍLIA**

**2021**

**PIETRO VICTOR MACIEL**

**UM ESTUDO SOBRE DADOS GOVERNAMENTAIS ABERTOS (DGA) NO  
CONTEXTO DO BRASIL COM O SUPORTE DE CIÊNCIA DE DADOS**

Relatório final de pesquisa de Iniciação Científica apresentado à Assessoria de Pós-Graduação e Pesquisa.

Orientação: Rodolfo Vasconcelos Correia Lima  
de Andrade

**BRASÍLIA**

**2022**

## **DEDICATÓRIA**

Nesta seção, gostaria de dedicar este artigo para minha mãe Simone, para meu irmão Lino e para o meu orientador Rodolfo Vasconcelos Correia Lima de Andrade.

## **AGRADECIMENTOS**

Para a confecção deste artigo, vários fatores foram importantes. Desse modo, gostaria de, nesta seção, agradecer minha família, sendo os principais apoiadores minha mãe Simone, meu irmão Lino e minha vó Eulália. Gostaria de agradecer meu orientador Rodolfo Vasconcelos Correia Lima de Andrade por me ajudar nesta jornada de conhecimento em uma nova área de interesse. Gostaria de prestar meus agradecimentos aos meus colegas Hericlys, Samuel, Júlia e Isabela, uma vez que me apoiaram em outros projetos durante o desenvolvimento deste artigo. Também, deixo minhas homenagens à Assessoria de Inicialização Científica do UniCEUB, que me deu uma oportunidade de publicar o primeiro de vários artigos que virão.

*“Não desista nas primeiras tentativas, a persistência é amiga da conquista.”*

*Bill Gates*

## RESUMO

O problema a ser estudado nesta pesquisa científica é sobre Dados Governamentais Abertos no contexto brasileiro com o suporte de Ciência de Dados. Desse modo, o artigo discorre sobre os conceitos de Dados Governamentais Abertos, quais são os princípios e leis essenciais para que um conjunto de dados seja considerado aberto. Além disso, é importante entender como países, como Estados Unidos, Canadá, Brasil e Estônia, lidam com o tema em suas nações. Desse modo, consegue-se traçar um comparativo entre eles e apresentar diferentes aplicações de Dados Governamentais Abertos. O artigo também discorre sobre os catálogos de Dados Abertos, utilizados para armazená-los em repositórios estruturados. Entretanto, é válido dizer que se trabalha com grande quantidade de dados. Logo, é importante filtrar as informações utilizando técnicas de Ciência de Dados. Desse modo, a pesquisa também traz explicações sobre Data Science e Big Data, Dados Grandes. Para o desenvolvimento deste artigo, utilizou-se uma abordagem qualitativa, focada em explicar, qualificar, o tema de Dados Abertos, e não trazer números sobre ele. Com isso, percebe-se que foi usada uma metodologia descritiva, preocupada em descrever o problema, em trazer novos conhecimentos ao leitor e tirar dúvidas sobre o conteúdo descrito. Para conseguir os dados, utilizou-se a metodologia bibliográfica e a laboratorial. A primeira foi usada para encontrar os autores que falam sobre Dados Governamentais Abertos. Já a segunda foi usada para buscar informações no repositório de dados do **Cenipa**, sobre ocorrências aéreas no Brasil. Não houve intervenção humana no desenvolvimento deste projeto. Foi proposto o desenvolvimento de uma solução prática que usa Dados Abertos. Para isso, mostra-se como foi a construção de uma **API**, Application Programming Interface, responsável por fazer o “meio-campo” entre o que o cliente digita no Facebook Messenger e os dados, que estão armazenados em um servidor da **AWS**, Amazon Web Service. Entendendo o fluxo do software desenvolvido, é possível construir um chatbot, Aerobot, que entrega ao cliente informações sobre ocorrências aéreas que acometeram aviões. É importante dizer que as camadas da solução encontram-se em ambiente gratuito, o que causa aumento no tempo de resposta. Ao fim consegue-se entender que, ainda que alguns conjuntos de dados estejam alocados em um catálogo de Dados Abertos, eles precisam concordar com os oito princípios levantados em 2007 por 30 defensores de Dados Abertos e a três leis, além da **LAI**, Lei de Acesso à Informação. Também, é possível apreender que esses repositórios estruturados ajudam no desenvolvimento de novas aplicações, como a desenvolvida nesta pesquisa, de Dados Abertos, disseminando-os para mais pessoas. Em suma, conclui-se que é necessário distribuir cada vez mais os dados, uma vez que isso aumenta a transparência com a população. E, quando se desejar realizar novos estudos, deve-se verificar se de fato são Dados Abertos os conjuntos de dados selecionados para a solução.

**Palavras-chave:** Dados Governamentais Abertos; Ciência de Dados; Aerobot; Cenipa.

## SUMÁRIO

1 INTRODUÇÃO	8
Objetivos	8
2 FUNDAMENTAÇÃO TEÓRICA	8
2.1 Dados abertos governamentais e transparência	8
2.2 Dados abertos governamentais no mundo	12
2.3 Repositório para catalogação de dados abertos	15
2.4 Grandes conjunto de dados	18
2.5 Ciência de dados	20
2.6 Análise de conjuntos de dados	24
2.6.1 Cenipa - Ocorrências aeronáuticas na aviação civil brasileir	24
2.6.1.1 Informações sobre o conjunto de dados do Cenipa	25
2.6.1.2 Análise do conjunto de dados do Cenipa	26
2.6.2 Série histórica de preços de combustíveis - Revenda	28
2.6.2.1 Informações sobre o conjunto de dados da ANP	28
2.6.2.2 Análise do conjunto de dados da ANP	29
2.6.3 Coronavírus (COVID-19)	31
2.6.3.1 Informações sobre o conjunto de dados do Ministério da Saúde	32
2.6.3.2 Análise do conjunto de dados do Ministério da Saúde	33
3 METODOLOGIA DE PESQUISA	35
3.1 Tipificação da pesquisa	35
3.2 Local da pesquisa	36
3.3 Objeto da pesquisa	36
3.4 Delimitação e universo da amostra	37
3.5 Instrumento de coleta de dados	37
3.6 Procedimentos metodológicos	37
4 RESULTADOS E DISCUSSÃO	39
5 CONSIDERAÇÕES FINAIS	40
REFERÊNCIAS	42
APÊNDICES	45

## **1. INTRODUÇÃO**

A presente pesquisa está inserida nos estudos relacionados a Dados Governamentais Abertos no contexto brasileiro com o suporte de Ciência de Dados. Assim, serão abordadas aplicações de Dados Abertos em outros países, qual a sua importância para aumentar a transparência socioeconômica e política. Percebe-se que esse tema é essencial, uma vez que, com a disseminação de dados abertos, a sociedade pode fiscalizar seu governo e questionar sobre possíveis incongruências encontradas. Desse modo, a pesquisa discorre sobre quais são os princípios dos dados abertos, os repositórios nos quais eles estão armazenados e como se consegue filtrar os dados. Ademais, serão analisados três conjuntos de dados, de modo que se possa verificar se são conjuntos de dados abertos ou não.

### **OBJETIVOS**

O objetivo principal desta pesquisa é construir um chatbot - chamado Aerobot - sobre ocorrências aéreas no Brasil. Para isso, o artigo busca apresentar o conceito, princípios/leis de dados abertos, a análise de 3 conjuntos de dados governamentais abertos disponíveis no repositório (catálogo) dados.gov.br e a importância de se utilizar dados abertos para fiscalização de informações obtidas em mídias sociais, televisivas.

## **2. FUNDAMENTAÇÃO TEÓRICA**

### **2.1 DADOS ABERTOS GOVERNAMENTAIS E TRANSPARÊNCIA**

Antes de definir dados abertos, não se pode ignorar o avanço da Internet e a onipresença das tecnologias de computação. É notório que os computadores já fazem parte da essência da sociedade e estão presentes no cotidiano de todas as pessoas. Mark Weiser, pai da Computação Ubíqua, define que as tecnologias mais profundas e duradouras são aquelas que desaparecem. Elas se dissipam nas coisas do dia a dia até se tornarem indistinguíveis (Mark Weiser, 1991). Com a facilidade de acesso aos meios de comunicação, mais dados são disponibilizados e podem ser interpretados pelas pessoas.



Entretanto, nem todo dado pode ser considerado aberto. Segundo a OKF (Open Knowledge Foundation) os dados são abertos quando qualquer pessoa tem acesso a eles e pode utilizá-los, reutilizá-los e distribuí-los, estando sujeitas a, no máximo, creditar a autoria ou compartilhar através da mesma licença (OKF, 2007). Através desse conceito, é perceptível que há uma maior interação entre as pessoas e as informações disponibilizadas na WEB, podendo elas ser compartilhadas. Assim, o conhecimento sobre determinado tema aumenta e debates construtivos entre as pessoas podem ser feitos.

Além do conceito dito acima, é importante citar quais os princípios dos dados abertos. Estes, definidos em 2007, na Califórnia, por um grupo de 30 defensores de dados abertos, foram divididos em 8 classificações, sendo elas: **a) Completos:** Todos os dados públicos precisam ser disponibilizados; **b) Primários:** Os dados devem ser publicados de forma bruta, conforme obtido pela fonte, evitando a transformação deles; **c) Atuais:** Eles devem ser publicados rapidamente, para que não perca seu valor e veracidade; **d) Acessíveis:** Devem ser distribuídos para o maior número de pessoas possível e para vários propósitos; **e) Interpretados por máquina:** Os dados devem ser disponibilizados de forma bruta e em vários formatos, facilitando a compreensão da máquina e de aplicações; **f) Acesso indiscriminado:** Os dados disponibilizados devem ser acessados por todos, sem haver nenhum tipo de distinção ou cadastro; **g) Não proprietário:** Nenhuma instituição deve ser dona ou controlar, de forma exclusiva, os dados abertos; **h) Sem licenças:** Nenhum dado aberto pode ser restringido por patentes, marcas, direitos autorais Como o próprio nome diz, são dados abertos. Entretanto, se houver necessidade de segurança ou privacidade, estas restrições devem ser regularizadas por estatutos.

Entretanto, para que o conceito de DAG (Dados Abertos Governamentais) seja considerado válido, o dado precisa ser regido por três leis (EAVES, 2009). Elas são:

- **Se o dado não for encontrado e indexado na web, ele não existe:** Nota-se, através dessa lei, que o dado precisa estar disponibilizado na rede, em repositórios conhecidos como catálogo de dados;
- **Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser aproveitado:** Através dessa lei, é visível a importância da distribuição de dados e informação. Para que eles possam ser lidos, aproveitados e republicados, considerando a existência de dispositivos heterogêneos, é necessário que tenham um padrão de formato. Assim, o dado deve ser publicado em JSON, XML, CSV ou RDF,

possibilitando a compreensão de todas as máquinas que usufruam dele (OPEN DATA HANDBOOK, 2010);

- **Se algum dispositivo legal não permitir sua replicação, ele não é útil:** O dado aberto não pode ser restringido por nenhum dispositivo legal, leis, normas, decretos. Se isso ocorrer, percebe-se sua inutilidade.

Princípios e leis foram levantados, mas algumas perguntas ainda precisam de respostas. Quais as utilidades de um dado aberto governamental? Qual a relação entre o cidadão e o governo? O país, no ano de 2020, está passando por um momento crítico com relação à pandemia do novo Coronavírus. O número de infectados e óbitos cresce a cada dia e todas essas informações são disseminadas por mídias tradicionais, como a imprensa. Contudo, é necessário ter um pensamento crítico para poder analisar os fatos e não só acreditar no que a mídia mostra. Esse controle não se restringe somente à situação da pandemia, podendo também ser feito para validar se os impostos pagos estão sendo utilizados de forma correta pelo governo, se ocorreu improbidade administrativa, se há corrupção, entre outras possibilidades. Nota-se o alto teor de controle social e fiscalização por parte da população. Com os Dados Abertos Governamentais, os cidadãos têm um maior controle da máquina estatal. Aliando-se esses dados aos conhecimentos das pessoas sobre tecnologia, é possível analisar os dados e gerar informações a partir deles. Sendo assim, é possível comparar as informações ditas pela imprensa e as geradas pela análise e pela tecnologia. Desse modo, consegue-se validar a veracidade da informação governamental.

Entretanto, a análise de Dados Abertos pode ocorrer a qualquer tempo. O cidadão tem um papel chave nesse processo. Isso ocorre, porque, se ele encontrar informações equivocadas ou se fizer uma previsão baseada nos dados abertos já existentes, o cidadão terá a oportunidade de alertar, fiscalizar e denunciar o Governo antes de que a máquina estatal publique essas informações. Isso aumenta a eficiência e interatividade entre cidadãos e Governo. Percebe-se o grande papel do cidadão na transparência e no desenvolvimento eficiente de sua nação.

Além de transparência, publicar dados abertos abre portas para que novos cruzamentos, novas interpretações e ferramentas de dados abertos sejam desenvolvidos (Dutra & Lopes, 2014). Com esses novos sistemas, fomenta-se a maior interação entre pessoas e governo. Muitas vezes, os cidadãos não têm conhecimento de que existem essas

ferramentas. Desse modo, ao desenvolver uma nova, a possibilidade de maior acesso e interesse por parte das pessoas aumenta. Com isso, benefícios, como: maior controle social e transparência; feedback sobre os conjuntos de dados; etc, surgem.

É notório que os cidadãos são o foco da disponibilização de dados abertos para a sociedade. Esse fato pode ser percebido, uma vez que as informações disponibilizadas pelos governos permitem que a inteligência coletiva crie melhores formas de trabalhar com elas do que os próprios governos poderiam fazer (SILVIA, 2010). Percebe-se, então, que essas informações podem ser utilizadas de acordo com a conveniência do interessado, de tal forma que elas possam ser cruzadas e combinadas para aumentar o valor dos dados (DINIZ, 2010). Desse modo, com a validação deles, as pessoas podem propor melhorias, participando coletivamente do desenvolvimento de seu país. Além disso, segundo a W3C, os governos devem incentivar os cidadãos a utilizar os dados abertos conforme sua necessidade e vontade, uma vez que não haverá valor na oferta deles se a população não reutilizá-los (DINIZ, 2010).

A transparência não deve ser um fim em si mesmo, mas um meio (ANGÉLICO, 2011). As atividades realizadas, assim como os dados, devem ser transparentes, para que todos possam obter conclusões construtivas e importantes. Para que os cidadãos lutem pelos seus direitos, os dados disponíveis devem ser relevantes para os interesses da população. Se os dados liberados não forem úteis para a sociedade em questão, não terão influência e não vão gerar questionamentos. Este é o conceito de grau de utilidade da transparência, onde os dados devem ser úteis e possibilitar inferências.

Independentemente da relevância dos dados abertos, deve-se incentivar a publicação deles. Com relação à importância, num momento de denúncia ou de análise, os dados, até então considerados “irrelevantes”, podem tornar-se essenciais. Sendo assim, deve-se buscar disponibilizá-los para um futuro acesso. Para incentivar a publicidade, alguns dispositivos legais foram criados, sendo eles a **Lei de Acesso à Informação - LAI Brasil**, criada em 2011; o **decreto Brasil (2012)** responsável pela regulamentação da LAI; o **decreto Brasil (2016)**, que estabelece políticas nacionais de abertura de dados no âmbito do executivo; e o **Portal Brasileiro de Dados Abertos**. Através das determinações especificadas nesses decretos, os órgãos têm o dever de publicar dados para fiscalização, fomentando a publicação e disponibilização.

O mundo atual está imerso em informação. Todas as atividades realizadas pelos governantes; os gastos entre conexões entre empresas e governo; impostos; investimentos devem ser amplamente estudadas e entendidas. Assim, é importante a obtenção de dados. Pensando nisso, os cidadãos devem assumir o papel de lutadores fiéis da sua nação. Não basta apenas observar. Deve-se procurar os dados, conseguir conhecimento e lutar por um país mais transparente e justo.

## **2.2 DADOS ABERTOS GOVERNAMENTAIS NO MUNDO**

A ideologia de Dados Abertos não é concentrada em um único país. Muitos buscam melhorar suas leis, derrubar a corrupção em prol de um país mais justo e correto. Tendo isso em mente, é necessário dizer que vários países adotam a abertura de informações governamentais, ainda que cada um busque aplicar esse conceito de distintas formas.

Sabe-se que todo novo conceito deve ser baseado em alguma doutrina ou artefato já existente. Essa situação ocorreu com o Brasil, que, com o objetivo de ter iniciativas governamentais relacionadas ao Governo Aberto, utilizou a Declaração de Governo Aberto, celebrada em setembro de 2011. Esse artefato considera que todas as pessoas anseiam por um governo mais transparente, exigindo por maior participação popular. Percebe-se que, com a população mais próxima do governo, é possível melhorar seus processos, ganhando eficiência na entrega dos serviços sociais, político-econômicos, entre outros.

Todos os países que assinam a Declaração de Governo Aberto entendem que eles devem ser mais transparentes e que as pessoas precisam participar dos diversos temas públicos. Ao aproximar os indivíduos dos seus governantes, cria-se a impressão de um país mais limpo, de um país que luta por seus habitantes. É necessário, então, escutar as exigências, propostas de melhorias, de inovações e de segurança. Para isso, a descentralização das informações se faz essencial para a Declaração descrita acima.

O Brasil também resolveu criar o Plano Nacional sobre Governo Aberto (PNGA - considerando que, ao fazer isso, o fortalecimento da democracia e da participação popular aumentaria. Assim, incrementar-se-ão a transparência, a prestação de serviços e a integração pública. Após a criação do PNGA e por experiências externas, conclui-se que é uma boa doutrina para aumentar a conexão governo-pessoa.

Outro fator importantíssimo para a busca de um Governo Aberto no Brasil é a Lei de Acesso à Informação, a **LAI**. Ela é um grande marco para a geração de cidadãos participativos e questionadores. Esses questionamentos e a abertura de dados ajudam na construção, desenvolvimento e destino da nação (HEIDEMANN, 2012).

Ao falar sobre Estados Unidos, é importante conectá-lo ao Brasil, uma vez que o país norte-americano é copresidente da Parceria para Governo Aberto (PGA) ao lado do governo brasileiro. Como citado anteriormente, cada nação aplica suas iniciativas de forma diferente. Nos EUA a diretriz adotada é composta por três princípios, sendo eles: transparência, participação e colaboração. A **transparência** se relaciona à importância de prestar contas à sociedade, disseminar informações. A **participação** é a entrega de insumos, ideias obtidas a partir da participação social. Por último, a **colaboração** é o ato de buscar relacionamentos com outros diferentes órgãos públicos e empresas privadas, alcançando mais eficiência e clareza.

Esse conceito de Governo Aberto é abordado por várias pessoas no mundo, inclusive líderes políticos, governamentais, entre outros. Isso pode ser percebido com a frase dita por Hillary Clinton, à época Secretária de Estado norte-americana. Segundo ela, o tratamento para o grande mal político (corrupção) é a abertura de informações. Com isso, as pessoas entenderão a situação pela qual o país está passando e poderão participar ativamente de mudanças e melhorias, destruindo a corrupção. Também citou que os governos devem ter interesse em apresentar as informações e ter as pessoas por perto. Com isso, todos saem ganhando, com uma nação mais limpa e forte político-socialmente (HILLARY, 2012).

Ultrapassando a fronteira norte-americana, outra potência internacional também tem os princípios de Dados Abertos enraizados em sua essência. Esta potência é o Canadá. A nação canadense divide Dados Abertos em três níveis. Eles são: **a) Dados Abertos:** liberação de dados governamentais para que cidadão, empresas privadas e ONG 's (Organizações Não Governamentais) possam utilizar de acordo com o seu interesse. Sendo assim, é possível inovar, criar aplicações que atendam às suas necessidades; **b) Informação Aberta:** deve haver uma continuidade na liberação de informações sobre atividades governamentais para os cidadãos canadenses; **c) Diálogo Aberto:** como pode-se perceber, deve ocorrer conversas entre governantes e cidadãos, podendo eles participar ativamente das decisões políticas, que tange ao interesse das pessoas (prioridades) daquela sociedade.

Como sempre, a tecnologia de informação está presente para facilitar e melhorar o acesso das pessoas a repositórios de dados abertos, a documentos WEB, entre outras funcionalidades. Sabendo que as pessoas estão mais ativas, correndo atrás de transparência pública e informações abertas, é possível criar uma ligação entre tecnologia e dados abertos. Os cidadãos desejam ter acesso e entender os problemas de sua nação, buscando opinar, debater e lutar pelo seu país, sendo protagonistas do seu futuro. Para esse intuito, a tecnologia cumpre seu papel, uma vez que repositórios, como o **Portal Brasileiro de Dados Abertos**, estão disponíveis na WEB, liberando o acesso para quem quer ser um agente ativo da nação (COUTINHO, 2012).

Existem pelo menos três fatores que impulsionam essa busca desenfreada pelo direito à informação. Esse aumento pode ser observado, como visto acima, em vários países e pode ser explicado pelas seguintes situações: **Participação ativa**, **Tecnologia** e **Globalização**. Antigamente poucas pessoas tinham direito a votar. Esse direito pertencia a homens com alto nível de estudo, ricos e inteligentes. Com o passar do tempo, a cidadania foi evoluindo até a liberação para que todos fossem dignos de votar. Percebe-se que, todavia, há limitações na **participação política**. Isso ocorre porque as pessoas votariam em seus candidatos e eles teriam liberdade para atender às suas promessas ou não, quando eleitos. Para aumentar a transparência e analisar as atividades desenvolvidas pelos governantes, é importante abrir o acesso aos dados. Os cidadãos anseiam por participar na administração de seus estados e de sua nação. Logo, somente permitir o voto para todos já não é o suficiente. Com a **tecnologia**, a sociedade terá maior acesso aos dados através de portais e entenderão o verdadeiro valor das informações. Por último, mas não menos importante, pode-se citar a **globalização** como fator determinante. Ao aplicar o conceito de aldeia global, em que as nações diminuíram suas fronteiras por causa das redes globais, é possível encontrar conexões internacionais. As pessoas se comunicam, podendo verificar o que está acontecendo em outras nações. Com isso, ao detectar-se uma melhoria ou ideia inovadora, é possível sugerir aos governantes e lutar por este novo direito. A comunicação, neste caso, é essencial para a mudança (TOBY, 2012).

Por último, mas não menos importante, pode-se citar a Estônia. Esta pequena nação do leste europeu é uma das maiores quando se fala em democracia digital. O que é **democracia digital**? A **democracia** acontece no momento em que se coloca a opinião da população como uma das engrenagens mais importantes para a criação de políticas públicas

voltadas às pessoas. Já o **digital**, é investir na democracia utilizando aplicativos, identificação digital e tecnologia. Ao falar sobre a Estônia, não é possível fugir dessa **rede de democracia**. Vale citar que essa proximidade entre Governo e cidadãos aumenta o conhecimento das pessoas, a transparência e agilidade da máquina estatal. Sendo assim, é possível ter um maior controle social e usar os Dados Abertos Governamentais como ferramenta de análise e eficiência governamental.

O berço para que a tecnologia entrasse na vida dos estonianos foi o X-Road. Este projeto, lançado na década de 90, ajudou na criação de um ambiente onde diversos sistemas de informação, sendo eles públicos ou particulares, pudessem gerar e trocar dados. É importante dizer que esse ambiente é padronizado e apresenta segurança para intercambiar informações. Isso é possível, uma vez que vários bancos de dados se conectam “por trás dos panos”, trabalhando em harmonia.

Com a espinha dorsal X-Road, foi possível gerar um modelo onde o Governo e empresas particulares pudessem criar aplicações cívicas. Elas são ferramentas, aplicativos, que utilizam dados abertos (presentes em uma mesma base de dados padronizada) para aumentar o bem-estar dos cidadãos, a transparência da máquina estatal e criar um país inovador e digitalmente conectado. É notório que nessa digitalização o maior beneficiado são os cidadãos..

Como forma de aplicação cívica estoniana, pode-se citar o **i-voting**. Essa ferramenta é utilizada como sistema de votação online. Ao invés de depositar seu voto em papel ou utilizar uma urna eletrônica, pode-se votar, em seus eleitores, utilizando quaisquer dispositivos conectados à Internet no mundo. Para isso, seu registro pessoal deve estar cadastrado. Durante as eleições, o usuário entra no sistema, utilizando um cartão de identidade (**RG Digital**) ou identificação móvel, e realiza seu voto. Para manter a privacidade do cidadão e o sigilo, a identidade de quem votou é retirada antes de chegar à Comissão Nacional Eleitoral.

Através da aplicação anterior, foi possível demonstrar como este país está conectado à tecnologia e como ela é utilizada em vários processos da sociedade e comunidades estonianas. Toda essa inovação ajuda na entrega de uma nação ágil, transparente e que leva seus cidadãos como peças essenciais para o avanço do país.

### **2.3 REPOSITÓRIOS PARA CATALOGAÇÃO DE DADOS ABERTOS**

É compreensível a importância de dados abertos. Contudo, de nada adianta ter os dados e não disponibilizá-los. Ao dividir o processo de abertura de dados em ciclos de abertura de dados, pode-se dizer que uma das etapas essenciais é a de publicação dos dados. Deve-se preocupar com essa etapa, uma vez que, para que a abertura ocorra de forma válida e para que os dados gerem valor para o usuário, sua publicação deve ser sustentável. Qual o resultado dessa etapa? Ela busca armazenar os dados em catálogos de dados, repositórios estruturados que permitem a leitura e redistribuição das informações contidas. Desse modo, uma interface é criada entre usuário e repositório (CÓRDOVA, 2013). Segundo a W3C, esses catálogos de dados são uma coleção com curadoria de metadados sobre conjuntos de dados (W3C, 2014).

Todos os dados abertos publicados na WEB estão presentes nesses repositórios. Com isso, é visível que eles são ferramentas responsáveis por gerir a publicação das informações. A partir do momento em que ocorre uma nova requisição, é responsabilidade do catálogo agir. Essas requisições podem ser a busca, inclusão e exclusão de dados e metadados (MIRANDA, 2013). Como exemplo para ferramenta de catálogo, pode-se citar a nação brasileira. Ela possui o Portal Brasileiro de Dados Abertos, ferramenta CKAN utilizada pelas pessoas para gerenciamento das atividades e dados advindos do governo nacional. Com ela, é possível ler e manipular suas informações para novas soluções tecnológicas.

Um dos grandes benefícios de uma ferramenta CKAN é a questão de ser um software livre. Além de não haver necessidade de investimento para utilizar esse software, é possível gerar catálogos de dados de alto nível que possibilitam publicação, armazenamento e manipulação. A ferramenta CKAN, criada pela Open Knowledge Foundation, hoje é gerenciada pela comunidade CKAN. Assim como outras tecnologias, como o Wordpress, ao ser gerenciada por uma comunidade, muitas dúvidas podem ser tiradas e aprendidas através da interação entre as pessoas daquele grupo. Isso é um fato importante, quando se busca aprender determinada ferramenta (CÓRDOVA, 2013). A partir de uma API, é possível verificar os dados através de gráficos, mapas, além de buscar em conjunto de dados geolocalizados. Contudo, o Brasil não é o único a utilizar a ferramenta CKAN em seu principal portal de dados. Várias nações, como: Austrália, Romênia, Argentina, entre outras, também utilizam.

Foi citado aqui o CKAN, contudo também existem outras ferramentas para catalogação de dados. O Junar Open Data Platform é um dos exemplos. Ele utiliza a ideia de



SAAS, onde o software é vendido como serviço. Não é necessário armazenar todo o software e ter a infraestrutura para executá-lo. Basta solicitar a uma empresa, que desenvolverá o sistema e o disponibilizará na nuvem. Desse modo, vários dispositivos diferentes terão acesso e a empresa contratante terá espaço livre para outras necessidades. Esse software Junar é localizado em **portais de dados abertos do Chile, Costa Rica, algumas cidades norte-americanas, entre outras** (JUNAR, 2014).

Pode-se citar o Socrata, que segue a mesma linha ideológica de SAAS. A principal diferença entre o Junar e o Socrata é que o segundo consegue construir visualizações mais elaboradas. Isso ocorre porque os dados dispostos no catálogo podem ser representados por mapas, gráficos, com formatação condicional, entre outras vantagens (MIRANDA, 2013). Alguns dos portais que utilizam esse sistema são: **portal oficial do Kenya, portal do Banco Mundial, Chicago, Nova York**, entre outras.

Há, também, um dos software mais recentes, o **OpenDataSoft**. Com ele, você pode especificar mais sua busca. Dentro dos conjuntos de dados, é possível filtrar através de parâmetros específicos. Quando se chega ao resultado esperado, automaticamente é gerado um código para acessar através de uma API. Este é um grande diferencial em relação aos anteriores. Esse software está presente principalmente na França, **em Brussels, Paris, Jornal L'Équipe, entre outros**.

Percebendo-se a evolução e as possibilidades dos outros sistemas de catalogação, é interessante optar pelo CKAN. Como já foi visto, ele é um software livre, ou seja, não há licenças nem é necessário pagar para utilizá-lo. Já os outros deve-se investir para seu uso. Nota-se que ocorre uma redução nos custos e a possibilidade de modificar (customizar) a partir do código fonte também existe. Ao utilizar o CKAN, abre-se uma vasta gama de possibilidades (GHOSH, 2006).

Percebe-se, até então, a importância de disponibilizar os dados, buscando redistribuição, inferências, conhecimento e melhorias sócio-governamentais. Entretanto, é interessante adentrar no tema relacionado a aplicativos. Através de um catálogo de dados, é possível criar aplicações que atendam determinados fins específicos, como: gerar novas visões e análises de dados. Essa maior exploração ajuda no aumento da transparência e da interação entre os cidadãos.

Como chegar ao desenvolvimento de um aplicativo? Muitas empresas particulares e públicas utilizam de hackathons para desenvolvimentos de novos sistemas (SILVEIRA, 2013).

Hackathons são competições nas quais várias pessoas de áreas distintas trabalham juntas na criação de uma nova solução tecnológica que atenda a determinados requisitos. A equipe pode ser formada por: programadores, ativistas, designers, analistas de sistemas e banco de dados, etc. Porém, “nem tudo são flores”. Muitos aplicativos construídos não vão para frente, uma vez que pode não haver público alvo ou seus sistemas não são produtos sustentáveis. Sendo assim, logo esses projetos são descartados (ORAM, 2011).

Neste momento de pandemia, deve-se investir em novas soluções que ajudem as pessoas a passar por essa situação. Com esse intuito, a agência de inovação e tecnologia da Estônia, em 2020, organizou um hackathon em apenas 6 horas, o **Hack the Crisis**. Como exemplo de solução gerada neste evento, pode-se citar o **Coronabot**. Essa ferramenta é um chatbot que responde perguntas sobre o coronavírus, contribuindo na diminuição do risco de contágio da doença. Para isso, dados abertos e verídicos devem ser utilizados nas soluções, ajudando na transparência e disseminação de informações corretas.

Como dito anteriormente, criar aplicativos baseados em dados abertos busca oferecer uma melhora na qualidade de vida dos cidadãos. A partir das aplicações, consegue-se reunir pessoas, criar debates, grupos de estudo e conversação para discutir sobre assuntos que sejam importantes para a sociedade. Percebe-se um aumento no interesse das pessoas por lutar por um Governo Aberto, onde a população tem seu direito de participar da gestão administrativa e governamental daquela nação (OLIVEIRA et al., 2013). Todo esse ambiente de conversa e conhecimento com relação aos dados abertos governamentais ajuda a fomentar a construção de novas comunidades e aplicações que mais se adaptem às comunidades (JOHNSON, 2010).

Ao criar uma nova solução tecnológica, deve-se levar em conta alguns requisitos. Sabe-se que, se não houver manutenção e cuidado, um software pode tornar-se obsoleto, “descontinuar”. Isso pode ocorrer porque os dados utilizados estão desatualizados, porque não há manutenção ou o projeto acabou. Ao levantar esses questionamentos, deve-se pensar e escolher muito bem o aplicativo que vai ajudar a resolver determinadas necessidades do indivíduo.

## 2.4 GRANDES CONJUNTOS DE DADOS

Dentro dos repositórios, existem dados coletados a partir das ações realizadas pelo governo. Ao considerar que um governo possui várias atividades e se divide em várias esferas governamentais, pode-se dizer que muitos dados são gerados diariamente. Eles também são atualizados, modificados, excluídos. Nota-se, então, que são grandes conjuntos de dados a serem manipulados, processados, avaliados para que possam gerar informação para quem os utiliza. Então Big Data é um pedaço do universo digital que inclui dados transacionais, dados armazenados, metadados e outros que estão presentes em arquivos enormes. Redes sociais, como Facebook e Instagram, também influenciam na expansão dos conjuntos de dados, uma vez que os usuários geram vários fluxos de dados continuamente (GANTZ, 2011).

Outro fator importante para esse grande volume de dados é a internet. Quando alguém deseja realizar uma análise correta, comparações com outras nações, procura-se conectar à rede para conseguir essas informações. Entretanto, todas as tecnologias empresariais, assim como os dispositivos móveis, geram bilhões de informações por dia. Sabe-se que nem todas são informações relevantes ou válidas para quem as busca. Desse modo, é necessária uma ferramenta para análise de dados, verificação de padrões e levantamento de informações importantes. Em um estudo realizado em 2012, foi possível constatar a existência de 2,8 zettabytes no mundo, o que equivale a  $2^{70}$  bytes de informação. A esse grande conjunto de dados, dá-se o nome Big Data.

A outra definição sobre Big Data cita que esse fenômeno surge porque os sistemas de gerenciamento de banco de dados (SGBD) tradicionais não são mais capazes de armazenar e analisar esses dados. Outra questão importante é dizer que os dados não estão mais estruturados, já que são obtidos de distintas fontes: e-mails, redes sociais, entre outras (MANYIKA, 2011).

Quando surgiu o conceito de Big Data, este termo pregava 3 V's com sendo essência da ideologia (LANEY, 2001). Eles são: **a) Volume** - o grande crescimento de dados não estruturados e a queda das tecnologias tradicionais de armazenamento e analíticas; **b) Velocidade** - de acordo com a incrementação dos recursos computacionais, com relação à produção, captura e processamento de dados, eles serão gerados de uma maneira mais rápida e veloz; **c) Variedade** - relaciona-se com as diferentes formas de obter os dados, através de sensores, vídeo/imagens, blogs, entre outras tecnologias.

Dados que não consigam ser interpretados, tratados e analisados em certo tempo, não servirão para criar informação e devem ser descartados pelas empresas e usuários.

Como qualquer processo, o Big Data deve entregar valor para quem está buscando. Logo, deve ser possível descobrir padrões, integrar dados que se encontravam em Data Warehouses distantes ou até em redes sociais, correios eletrônicos, no YouTube, Twitter, entre outras tecnologias. Assim, é possível utilizar o melhor da ferramenta e gerar realmente o que as pessoas estão buscando. Data Warehouse é uma grande base de dados que centraliza e organiza os dados por assuntos integrados. Dessa forma, quem desejar utilizar esse repositório conseguirá relatórios complexos, ajudando nas tomadas de decisão e melhorando a transparência governamental. Assim, diz-se que Data Warehouse é um repositório com dados integrados (SHAMS e FARISHTA, 2001).

As tecnologias utilizadas para Big Data podem ser analisadas por dois aspectos diferentes: **a) Analítica:** relacionado às tecnologias usadas para análise de dados. Dentre elas, pode-se citar o Hadoop e MapReduce; **b) Infraestrutura:** aspecto relacionado com a tecnologia de armazenamento e processamento dos dados. Dentre essas ferramentas, pode-se citar o NoSQL (KERNOCHAN, 2011).

## 2.5 CIÊNCIA DE DADOS

Ao falar sobre Big Data, pode-se concluir que ela representa uma base complexa constituída por uma imensa quantidade de dados. Cabe à ciência de dados analisar e extrair informações úteis de dados complexos, dinâmicos, distintos e distribuídos (BUGNION; MANIVANNAN; NICOLAS, 2017). As características (domínios) da área de Data Science dividem-se em três: **Programação de Computadores, Estatística e Matemática e Domínio do Conhecimento**. A partir destas características, é possível fazer três conjecturas:

- Todos os cientistas de dados devem ter domínio da Ciência da Computação. Sabe-se que os dados, para serem coletados, armazenados, manipulados, devem estar presentes nos computadores. Todas essas etapas são realizadas por sistemas computacionais, softwares. Desse modo, é imprescindível o conhecimento sobre tecnologia. Com isso, os dados podem ser analisados e transformados em informações úteis. Essa conjectura ajuda a promover a Curadoria Digital e a implementação de algoritmos de Aprendizado de Máquina e de interfaces

necessárias para Visualização da Informação, gráficos, mapas, entre outras representações;

- A segunda conjectura considera a **Matemática e Estatística** como áreas essenciais para a Ciência de Dados. Sabe-se que, a partir da análise e transformação, informações surgirão. Cria-se uma necessidade de poder validar e interpretar essas informações estatisticamente. Além de entender os algoritmos para Aprendizado de Máquina, os “data scientists” devem conseguir interpretar. Percebe-se a grande utilidade dessas áreas exatas na Ciência de Dados. Essa interpretação pode ser facilitada a partir do momento em que as interfaces de visualização da Informação são feitas, uma vez que é possível representar graficamente as informações.
- Deve-se, também, considerar que o Conhecimento do Domínio deve estar presente para criar uma solução tecnológica que atenda aos clientes, às pessoas. Por quê? Em qualquer área, caso não seja compreendida a área em que se está atuando, o resultado será equivocado. O mesmo ocorre com a Ciência de Dados. É necessário gerar insumos (informações) importantes para ajudar no processo final, a tomada de decisão, por exemplo. Se a solução não obtiver esse resultado, é necessário revisá-la e buscar atender ao que é pedido.

Ao falar sobre coleta de dados, é essencial citar o motivo para este processo. Para tomar uma decisão, tanto em âmbito social quanto em âmbito organizacional, dados devem ser obtidos. Através de indícios e padrões verificados, é possível analisar e escolher a melhor forma de procedência, decidir. Esse é o padrão adotado pelo Big Data (Provost; Fawcett, 2013). Para coletar, analisar e extrair informações relevantes, papel da ciência de dados, normalmente utiliza-se soluções (ferramentas) computacionais. As principais ferramentas utilizadas são algoritmos de Aprendizado de Máquina. Mas, conceitualmente, o que significa Aprendizado de Máquina. Ao falar sobre este conceito, é possível defini-la como uma subárea da Inteligência Artificial (**IA**) utilizada para extrair informações valiosas a partir de uma grande (volumosa) base de dados (Blum; Hopcroft; Kannan, 2018). Essa ferramenta utiliza métodos estatísticos para auxiliar neste processo anteriormente citados. Sabendo destes conceitos, alguns métodos computacionais podem ser levantados. Eles são:

- **Redes Neurais Artificiais:** Como é possível perceber através do nome, neural se relaciona com os neurônios, cuja localização é no cérebro humano. Desse modo, é possível abstrair que esse modelo busca imitar as funções deste órgão. Assim como o cérebro, uma rede neural recebe estímulos (sinais de entrada), processa-os e gera uma saída (Munakata, 2008).
- **Algoritmos Genéticos:** Ao falar sobre Algoritmos Genéticos, é possível fazer uma analogia com a teoria da evolução das espécies, de Charles Darwin (Munakata, 2008). Essa teoria diz que somente os seres que mais conseguem adequar-se ao ambiente (ecossistema) sobrevivem. A partir dessa proposição, pode-se dizer que o algoritmo genético deve selecionar aquelas soluções computacionais (**seres vivos**) que mais se adequem ao ambiente computacional e que sejam mais aptas a resolver determinado problema; a reprodução das melhores soluções; e a ocorrência ocasional de mutação sobre as soluções. Através dessa analogia, pode-se concluir que o papel de um Algoritmo Genético é encontrar uma solução ótima (que atenda aos requisitos) dentre várias existentes.
- **Inteligência Coletiva:** Esse algoritmo nasceu no contexto de sistemas autônomos que se baseavam na coletividade e auto-organização de agentes simples (Tarasewich; McMullen, 2002). O algoritmo de Inteligência Coletiva busca compreender o comportamento coletivo para que determinada tarefa global seja realizada. Pode-se dizer que a Inteligência Coletiva se inspira na biologia. Como assim? Para responder essa pergunta, o seguinte exemplo pode ser dado: **a formiga deve trabalhar em equipe (conhecimento coletivo) para conseguir levar seus alimentos para a colônia, assim como as abelhas devem coletar o néctar e levar para a colméia.** O algoritmo deve perceber esses comportamentos para que possam ser feitas otimizações. Estes algoritmos são muito utilizados para mineração de dados ou descoberta de conhecimento em bases de dados (Abraham; Guo; Liu, 2006).

Independentemente do modelo computacional escolhido, sete passos devem ser seguidos para uma solução de Data Science, **Ciência de Dados** (BUGNION; MANIVANNAN; NICOLAS, 2017). Eles são:

- **Obtenção de dados:** É a fase em que dados primários são avaliados e coletados. Esses dados brutos são encontrados em arquivos de texto; em sensores; em imagens de câmeras; na própria internet; etc;
- **Ingestão de dados:** Deve-se considerar que os dados obtidos são de origem diferente e possuem formatos distintos. Sabendo disso, eles devem ser transformados e colocados em uma base de dados principal, centralizada e organizada. Isso ocorre uma vez que se tem os dados em um lugar específico, o esforço é menor e o processamento torna-se mais rápido
- **Exploração de dados:** Ao explorar os dados disponíveis na base, deve-se verificá-los, estudá-los e dar pareceres preliminares relacionados à informação solicitada. Desse modo, consegue-se montar um roteiro de como relacionar os dados brutos (primários) de forma a gerar uma informação relevante;
- **Definição dos parâmetros:** Neste momento, algumas atividades são realizadas. Estas são: conversão dos dados de acordo com os requisitos do algoritmo de aprendizado, transformação dos dados para entendimento dos seres humanos, definição dos critérios de parada do algoritmo, nível de maturidade e confiabilidade esperado da resposta final, entre outras definições;
- **Implementação do modelo:** A partir dos dados de entrada e de saída e da definição dos parâmetros, deve-se dizer que, com o uso de uma ferramenta de Aprendizado de Máquina, é possível estabelecer modelos para posterior aplicação. Utilizando esse processo de forma iterativa, consegue-se utilizar estratégias e fazer testes no algoritmo, buscando melhorias e atualizações nos parâmetros já definidos. Com isso, um modelo mais abstrato (considerando os dados presentes) pode ser desenvolvido;
- **Utilização do modelo:** Após a definição de um modelo, deve-se utilizá-lo. O motivo disso é que se sabe que esse algoritmo estará presente em ambientes reais, situações reais. Sendo assim, deve-se usar o modelo para verificar o poder de inferência de informações relevantes em um ambiente de produção. Se esse poder for confirmado, pode-se aplicar o modelo em tarefas intensivas de conhecimento;
- **Tomada de decisão:** Após a análise feita através do conhecimento pertencente às pessoas e das informações apresentadas pelo software, uma decisão pode ser tomada. Diz-se que ela está embasada, uma vez que se baseia em uma informação real gerada pelo sistema. É importante dizer que utilizar gráficos e relatórios para

apresentação e visualização das informações. A razão disso é que, gerando gráficos mais customizados e claros, é possível ajudar as pessoas a ter insights mais claros e convincentes. Dessa forma, a decisão é facilitada.

Todo esse processo (atividades) anteriormente dito ocorre de forma iterativa, ou seja, através de ciclos contínuos. Sabe-se que o papel principal de utilizar dados analisados e extraídos do algoritmo de Aprendizado de Máquina é ajudar na tomada de decisão das pessoas. Desse modo, pode-se iniciar com uma premissa mais geral, refina-la de acordo com o avanço, teste e utilização de outros algoritmos de aprendizado. Percebe-se, com isso, a possibilidade de encontrar soluções (modelos) mais úteis para a tomada de decisão, sendo elas obtidas, iterativamente, através de um grande volume de dados.

## **2.6 ANÁLISE DE CONJUNTOS DE DADOS**

Com os conceitos em mente, buscou-se analisar alguns conjuntos de dados presentes no Portal Brasileiro de Dados Abertos. O objetivo dessa análise é verificar se os conjuntos de dados escolhidos para essa pesquisa são realmente Dados Abertos Governamentais. Para isso aplicar-se-ão os conceitos de Dados Abertos, os princípios e leis já relatados neste projeto. Os conjuntos de dados escolhidos são: **Cenipa - Ocorrências Aeronáuticas na Aviação Civil Brasileira; Série Histórica de Preços de Combustíveis - Revenda; e Coronavírus (COVID-19).**

### **2.6.1 CENIPA - OCORRÊNCIAS AERONÁUTICAS NA AVIAÇÃO CIVIL BRASILEIRA**

A base de ocorrências escolhida acima é gerida pelo Centro de Investigação e Prevenção de Acidentes Aeronáuticos, o famoso **CENIPA**. Seu objetivo é demonstrar dados referentes a ocorrências aéreas (acidentes e incidentes) que aconteceram nos últimos 10 anos na nação brasileira.

Com relação aos dados encontrados neste conjunto, o usuário terá acesso a informações relacionadas ao incidente, como: **1. Quais foram as aeronaves envolvidas?; 2. Onde aconteceu?, 3. Em qual dia e horário?.** Além das respostas referentes a essas perguntas, o conjunto de dados apresentará algumas informações taxonômicas das



investigação de acidentes, **AIG**. É importante dizer que dados relacionados às pessoas físicas e jurídicas presentes no acidente são resguardados, privados, conforme previsto pela Lei de Acesso à Informação (Lei nº 12.527, de 18 de novembro de 2011).

Mas de onde surgiram as informações presentes nesta base de dados? Elas foram obtidas no formulário CENIPA-05 (Ficha de Notificação de Ocorrências Aeronáuticas) e solidificadas com os relatórios gerados pelas investigações publicadas. A outra maneira de conseguir enxergar esses dados é acessar o **PAINEL SIPAER**, encontrado na página do CENIPA.

Dados originados nos Programas de Gerenciamento de Prevenção, como: **Risco de Fauna, Emissão de Raio Laser e Risco Baloeiro**, apesar de serem geridos pelo CENIPA, não se encontram neste conjunto de dados. Isso acontece porque, enquanto os dados relativos a esse programa têm maior foco no gerenciamento de risco, o CENIPA preocupa-se mais em publicar dados relativos a ocorrências aéreas.

#### 2.6.1.1 INFORMAÇÕES ADICIONAIS SOBRE O CONJUNTO DE DADOS DO CENIPA

O Quadro 1 apresenta algumas informações adicionais referentes ao conjunto de dados de ocorrências aeronáuticas gerido pelo CENIPA.

Quadro 1: Informações Extras do CENIPA

<b>CAMPO</b>	<b>VALOR</b>
<b>Fonte</b>	<a href="https://www2.fab.mil.br/cenipa/">https://www2.fab.mil.br/cenipa/</a>
<b>Autor</b>	Centro de Investigação e Prevenção de Acidentes Aeronáuticos
<b>Mantenedor</b>	Centro de Investigação e Prevenção de Acidentes Aeronáuticos
<b>Versão</b>	1.3
<b>Última Atualização</b>	27 de novembro de 2020, 11:16 (UTC-03:00)

<b>Criado</b>	1 de junho de 2015, 15:37 (UTC- 03:00)
<b>Cobertura Geográfica</b>	Brasil
<b>Cobertura Temporal</b>	2010 a 2019
<b>Frequência de Atualização</b>	Anual
<b>Granularidade Geográfica</b>	Aeródromo
<b>Granularidade Temporal</b>	Hora:Minuto
<b>VCGE</b>	Aeronáutica [ <a href="http://vocab.e.gov.br/2011/03/vcge#aeronáutica">http://vocab.e.gov.br/2011/03/vcge#aeronáutica</a> ], Transporte Aéreo [ <a href="http://vocab.e.gov.br/2011/03/vcge#transporte-aéreo">http://vocab.e.gov.br/2011/03/vcge#transporte-aéreo</a> ]

Fonte: Portal Brasileiro de Dados Abertos

### 2.6.1.2 ANÁLISE DO CONJUNTO DE DADOS DO CENIPA

A seguir serão apresentados questionamentos para verificar se o conjunto de dados do Cenipa pode ser considerado Dados Abertos.

#### 1 - O conjunto de dados do Cenipa é completo?

Antes da análise, é necessário entender o que a organização Cenipa deseja mostrar ao desenvolver um conjunto de dados sobre ocorrências aéreas em solo brasileiro. Como observação válida, é interessante dizer que a aviação torna-se melhor e mais segura a cada dia que se passa. O motivo dessa evolução é porque há transparência e dados, protocolos são disponibilizados. Desse modo, tanto as empresas aéreas quanto as pessoas têm acesso a eles e podem aprender e atualizar a aviação civil. Verificando o conjunto de dados e sabendo que a aviação brasileira preza pela transparência, é notório que os dados disponibilizados pelo Cenipa são e devem ser completos. Com relação ao que este conjunto propõe, consegue-se encontrar, na base, dados referentes às aeronaves; aos códigos delas; ao tipo de operação das aeronaves; ao país de fabricação e à empresa; ao número de assentos delas; ao segmento, como táxi aéreo e carga. Pensando no evento, pode-se encontrar informações

relacionadas ao código e ao tipo da ocorrência, ao local, latitude e longitude do acidente, ao status da investigação, ao número de recomendações pós-acidente, entre outras. Através disso, pode-se dizer que este conjunto de dados é completo.

## **2 O conjunto de dados do Cenipa é primário?**

Com relação aos dados presentes nas tabelas do conjunto, consegue-se verificar que eles se encontram em formato bruto. Pensando-se em banco de dados, encontra-se apenas coluna e linha, atributo e valor dentro de planilhas. Evita-se a transformação, cruzamento e comparação com outros dados.

## **3 - O conjunto de dados do Cenipa é atual?**

Com os metadados expostos no Quadro 1, percebe-se que a última atualização feita, considerando que a base foi criada em 2015, ocorreu no dia 27 de novembro de 2020. Sendo este projeto desenvolvido no ano de 2020 e estando no mês de novembro, é possível afirmar que os dados são atuais. Desse modo, são dados úteis que possuem valor.

## **4 - O conjunto de dados do Cenipa é acessível?**

Por estar disponibilizado em um repositório de dados importante no Brasil, o Portal Brasileiro de Dados Abertos, ser de livre acesso e distribuição para vários propósitos, diz-se que é um conjunto de dados acessível.

## **5 - O conjunto de dados do Cenipa é interpretado por máquina?**

Analisando as tabelas e os dados relacionados a ocorrências aéreas, percebe-se que estão dispersos em vários formatos, como: **CSV, HTML, PDF, formato de imagem**. Sabendo das recomendações de formatos e do conceito de interpretação da máquina, pode-se dizer que é um conjunto de dados interpretado por máquina.

## **6 - O conjunto de dados do Cenipa tem acesso indiscriminado?**

Os dados postados pelo Cenipa possuem acesso indiscriminado, uma vez que todas as pessoas têm acesso, independente de características pessoais, físicas, financeiras, entre outras. O acesso é feito por pessoas que têm interesse em saber sobre ocorrências aéreas e estudar sobre o assunto.

### **7 - O conjunto de dados do Cenipa tem propriedade?**

Propriedade é quando alguma empresa ou organização tem acesso exclusivo ou é dono dos dados, da informação. Com este conceito “em mãos”, pode-se dizer que este conjunto de dados do Cenipa não tem propriedade. A diferença é que quem alimenta as tabelas e o conjunto é o Cenipa.

### **8 - O conjunto de dados do Cenipa tem licenças?**

A única licença que este conjunto de dados possui é a Open Database License, **ODbl**. Qual sua definição? É um acordo feito com a copyleft que permite que qualquer usuário que acesse o banco de dados faça modificações e tenha liberdade para compartilhar. Sendo assim, é uma licença que só garante o conceito de Dados Abertos Governamentais.

### **9 - O conjunto de dados do Cenipa pode ser replicado?**

Conforme visto no conceito da licença ODbl, o conjunto do Cenipa é, sim, replicável.

### **10 - Com todas as respostas acima, o conjunto de dados do Cenipa sobre ocorrências aéreas é considerado Dado Aberto?**

Com a validade de todos os questionamentos acima, conclui-se que este conjunto de dados é, sim, aberto.

## **2.6.2 SÉRIE HISTÓRICA DE PREÇOS DE COMBUSTÍVEIS - REVENDA**

Como é possível perceber através do nome do conjunto de dados, ele trata sobre os valores da venda de combustível em postos de gasolina. O objetivo desta base de dados é apresentar o histórico de valores e o quanto ele foi aumentando com o decorrer dos anos. Alguns dos dados persistidos na base são: regiões, anos, preços, combustível, postos, entre outros. A partir desses dados, abre-se um leque de possibilidades para soluções relacionadas a essas informações.

### 2.6.2.1 INFORMAÇÕES ADICIONAIS SOBRE O CONJUNTO DE DADOS DA ANP

No Quadro 2, apresentado a seguir, serão exibidos metadados importantes relacionados ao conjunto de dados referente ao histórico do preço de combustível no Brasil.

Quadro 2: Informações Extras da ANP

<b>CAMPO</b>	<b>VALOR</b>
<b>Fonte</b>	<a href="http://www.anp.gov.br/dados-abertos-anp">http://www.anp.gov.br/dados-abertos-anp</a>
<b>Autor</b>	Agência Nacional do Petróleo, Gás Natural e Biocombustível
<b>Mantenedor</b>	<a href="http://anp.gov.br/fale-conosco">anp.gov.br/fale-conosco</a>
<b>Versão</b>	1.0
<b>Última Atualização</b>	29 de novembro de 2020, 14:41 (UTC-03:00)
<b>Criado</b>	21 de agosto de 2018, 12:35 (UTC- 03:00)
<b>Escopo Geopolítico</b>	Nacional
<b>Fonte</b>	Sistema SLP
<b>Frequência de Atualização</b>	Mensal
<b>Idioma</b>	Português
<b>Órgão - Esfera</b>	Federal
<b>Órgão - Poder</b>	Executivo

Fonte: Portal Brasileiro de Dados Abertos

### 2.6.2.2 ANÁLISE DO CONJUNTO DE DADOS DA ANP

A seguir serão apresentados questionamentos para verificar se o conjunto de dados da ANP pode ser considerado Dados Abertos.

### **1 - O conjunto de dados da ANP é completo?**

Todos os dados públicos relacionados ao histórico de preços dos combustíveis no Brasil são abordados neste conjunto de dados. Percebe-se isso no momento em que são exibidos dados, como: preço do combustível nas regiões, regiões, anos, municípios, unidades revendedoras, etc. Assim, todos os dados públicos sobre este assunto são encontrados nas tabelas desta base de dados.

### **2 - O conjunto de dados da ANP é primário?**

Os dados se encontram em formato bruto e não há cruzamentos e transformações entre eles. Isso facilita na hora da construção de uma nova aplicação que utilize este conjunto de dados, porque o desenvolvedor terá mais possibilidades de atender ao seu propósito, adaptando os dados para o seu interesse. Os dados são encontrados em planilhas, com coluna e valor, mantendo um estado de dado bruto.

### **3 - O conjunto de dados da ANP é atual?**

Com os metadados expostos no Quadro 2, percebe-se que a última atualização feita, considerando que a base foi criada em 2018, ocorreu no dia 29 de novembro de 2020. Sendo este projeto desenvolvido no ano de 2020 e estando no mês de novembro, é possível afirmar que os dados são atuais. Desse modo, são dados úteis que possuem valor.

### **4 - O conjunto de dados da ANP é acessível?**

Por estar disponibilizado em um repositório de dados importante no Brasil, o Portal Brasileiro de Dados Abertos, ser de livre acesso e distribuição para vários propósitos, diz-se que é um conjunto de dados acessível.

### **5 - O conjunto de dados da ANP é interpretado por máquina?**

Analisando as tabelas e os dados relacionados a ocorrências aéreas, percebe-se que estão dispersos em formatos, como: **CSV** e **PDF**. Sabendo das recomendações de formatos e do conceito de interpretação da máquina, pode-se dizer que é um conjunto de dados interpretado por máquina.

#### **6 - O conjunto de dados da ANP tem acesso indiscriminado?**

Os dados persistidos pela ANP possuem acesso indiscriminado, uma vez que todas as pessoas têm acesso, independente de características pessoais, físicas, financeiras, entre outras. O acesso é feito por pessoas que têm interesse em saber sobre o histórico de evolução dos preços dos combustíveis.

#### **7 - O conjunto de dados da ANP tem propriedade?**

Propriedade é quando alguma empresa ou organização tem acesso exclusivo ou é dono dos dados, da informação. Com este conceito “em mãos”, pode-se dizer que este conjunto de dados da ANP não tem propriedade. A diferença é que quem alimenta as tabelas e o conjunto é a ANP.

#### **8 - O conjunto de dados da ANP tem licenças?**

A única licença que este conjunto de dados possui é a **Creative Commons Attribution**. Qual sua definição? São várias licenças que autorizam a distribuição, de forma gratuita, de uma obra (dados) com direitos autorais protegendo-os. Ela é utilizada para que as pessoas possam compartilhar, usar e construir soluções utilizando esses dados previamente disponibilizados.

#### **9 - O conjunto de dados da ANP pode ser replicado?**

Conforme visto na definição da licença Creative Commons, o conjunto da ANP é, sim, replicável.

#### **10 - Com todas as respostas acima, o conjunto de dados da ANP é considerado Dado Aberto?**

Com a validade de todos os questionamentos acima, conclui-se que este conjunto de dados é, sim, aberto.

### **2.6.3 CORONAVÍRUS (COVID-19)**

Foi abordado este conjunto de dados, uma vez que infelizmente ele está muito presente no momento em que a sociedade mundial se encontra. O mundo está enfrentando uma luta desenfreada contra o avanço deste vírus. Uma boa ferramenta para verificar se o número de casos está controlado ou aumentando é a distribuição de um conjunto de dados relacionado ao coronavírus. Por esse motivo, escolheu-se a base de dados do Ministério da Saúde.

O que vem a ser o coronavírus? É uma família de vírus responsável por desenvolver várias infecções respiratórias nas pessoas. Foi descoberto este novo agente em 31/12/2019 porque vários casos vinham sendo repercutidos na China. A doença provocada por este vírus é a COVID-19. Servindo de informação base para este conjunto, os primeiros vírus encontrados em seres humanos foram isolados em 1937. Contudo, somente em 1965, o vírus foi chamado por coronavírus. Isso ocorreu por causa da verificação microscópica, onde foi possível enxergar um vírus em formato de coroa.

Grande parte das pessoas, durante sua vida, é infectada por coronavírus comuns. Como forma de sanar curiosidades, pode-se dizer que as crianças são as mais propensas a serem infectadas. Os coronavírus mais comuns em humanos são: **alpha coronavírus 229E** e **NL63** e **beta coronavírus OC43**.

### 2.6.3.1 INFORMAÇÕES SOBRE O CONJUNTO DE DADOS DO MINISTÉRIO DA SAÚDE

No Quadro 3, apresentado a seguir, serão exibidos metadados importantes relacionados ao conjunto de dados referente ao número de vítimas do Coronavírus em território nacional.

Quadro 3: Informações Extras do Ministério da Saúde

CAMPO	VALOR
Fonte	<a href="https://coronavirus.saude.gov.br/">https://coronavirus.saude.gov.br/</a>



<b>Autor</b>	Ministério da Saúde
<b>Mantenedor</b>	DataSus
<b>Última Atualização</b>	7 de abril de 2020, 10:29 (UTC-03:00)
<b>Criado</b>	6 de abril de 2020, 16:00 (UTC- 03:00)
<b>Abrangência</b>	Nacional
<b>Periodicidade</b>	Diária

Fonte: Portal Brasileiro de Dados Abertos

### 2.6.3.2 ANÁLISE DO CONJUNTO DE DADOS DO MINISTÉRIO DA SAÚDE

A seguir serão apresentados questionamentos para verificar se o conjunto de dados do Ministério da Saúde pode ser considerado Dados Abertos.

#### 1 - O conjunto de dados do Ministério da Saúde é completo?

Ao tentar acessar a planilha de dados do coronavírus disponibilizada pelo Ministério da Saúde, encontra-se um erro. O arquivo CSV não foi encontrado, sendo assim não é possível saber se os dados são completos ou não. Outro argumento que defende isso é a última atualização do arquivo. Sabendo que o coronavírus é um problema crescente, dados referentes ao mês de abril já não têm o mesmo valor. Isso reforça o argumento de que o conjunto não é completo.

#### 2 - O conjunto de dados do Ministério da Saúde é primário?

Apesar de não ser possível acessar a planilha CSV disponibilizada pelo **MS**, Ministério da Saúde, ao entrar no site, também é possível encontrar vários gráficos relacionados aos números de Covid no Brasil. Para a modelagem de gráficos, é necessário fazer comparações, cruzamentos. Desse modo, não está presente o dado bruto, o dado primário. Eles já foram adaptados para um determinado problema.

#### 3 - O conjunto de dados do Ministério da Saúde é atual?

Conforme foi dito anteriormente, a última atualização foi feita no dia 7 de abril de 2020. Desde então, vários dados mudaram e aumentaram. Sendo assim, as informações presentes nessa “planilha” são inexistentes.

#### **4 - O conjunto de dados do Ministério da Saúde é acessível?**

Apesar de aparecer dentro do principal repositório de dados abertos do Brasil, o Portal Brasileiro de Dados Abertos, a planilha CSV que apresentaria dados brutos não pode ser acessada. Sendo assim, não se pode considerar este conjunto de dados abertos acessível.

#### **5 - O conjunto de dados do Ministério da Saúde é interpretado por máquina?**

Ainda que não seja possível encontrar a planilha, ela utiliza o formato CSV, que a torna interpretada por máquina.

#### **6 - O conjunto de dados do Ministério da Saúde tem acesso indiscriminado?**

Os dados persistidos pelo Ministério da Saúde possuem acesso indiscriminado, uma vez que características pessoais, físicas, financeiras, entre outras não são levadas em conta. Contudo, como não é possível acessar a planilha com os dados brutos, cria-se um limite.

#### **7 - O conjunto de dados do Ministério da Saúde tem propriedade?**

Propriedade é quando alguma empresa ou organização tem acesso exclusivo ou é dono dos dados, da informação. Com este conceito “em mãos”, pode-se dizer que este conjunto de dados do Ministério da Saúde não tem propriedade. A diferença é que quem alimenta as tabelas e o conjunto é o DataSus.

#### **8 - O conjunto de dados do Ministério da Saúde tem licenças?**

Possui a licença **Creative Commons Attribution**, utilizada para que as pessoas possam compartilhar, usar e construir soluções utilizando esses dados previamente disponibilizados de forma gratuita.

#### **9 - O conjunto de dados do Ministério da Saúde pode ser replicado?**

Conforme visto na definição da licença Creative Commons, o conjunto do Ministério da Saúde também é replicável.

## **10 - Com todas as respostas acima, o conjunto de dados do Ministério da Saúde é considerado Dado Aberto?**

Por não seguir todos os princípios e leis sobre dados abertos (conceitos já discutidos neste trabalho) pode-se dizer que não é um conjunto de dados abertos.

### **3. METODOLOGIA DE PESQUISA**

#### **3.1 TIPIFICAÇÃO DA PESQUISA**

Neste projeto, foi utilizada uma abordagem qualitativa, em que se busca aproximar de autores que estudam sobre o tema de Dados Abertos Governamentais. São apresentados conceitos, países que utilizam-nos como forma de transparência para a população, além de uma análise apurada em cima de três bases de dados consideradas abertas. Percebe-se que o objetivo é qualificar um tema, e não trazer números sobre ele. É procurar entendê-lo, estudá-lo e apresentar essas informações ao leitor.

Entendendo a abordagem, compreende-se que, com relação aos objetivos da pesquisa, foi utilizada uma metodologia descritiva. A pesquisa concentrou-se em cima de um tema e em que autores publicaram sobre ele. O objetivo deste artigo é discorrer sobre Dados Abertos Governamentais, buscando esclarecer quaisquer dúvidas ou falta de conhecimento pertinentes ao conteúdo apresentado. O artigo tem o interesse em gerar curiosidade nos leitores, alertando-os sobre um tema essencial para uma sociedade mais transparente quanto às suas ações políticas e socioeconômicas.

Com relação aos procedimentos para buscar os dados, foram utilizadas duas metodologias, tanto bibliográfica quanto laboratorial. A primeira foi utilizada para conseguir dados essenciais em artigos e na Internet para embasar a pesquisa, sendo esta metodologia muito importante para o desenvolvimento deste artigo. Além disso, desenvolveu-se uma solução que utiliza Dados Abertos Governamentais, sendo essas outras informações adquiridas com a metodologia laboratorial. Nesta segunda metodologia, utilizou-se um

ambiente computacional para acessar o repositório *dados.gov.br* e buscar as bases de dados necessárias para a aplicação.

Quanto à natureza desta pesquisa científica, ela é classificada como pesquisa aplicada. Isso ocorre uma vez que o seu objetivo, além de apresentar os conceitos e embasar o tema, é gerar uma aplicação prática com o objetivo de solucionar problemas específicos. Logo, busca-se aplicar os conceitos apresentados.

### **3.2 LOCAL DA PESQUISA**

Com relação ao local de realização da pesquisa, foi utilizado um ambiente computacional. Não foi necessário ter contato com pessoas físicas. Foi importante apenas o uso de equipamentos tecnológicos (computador, notebook) com acesso à Internet. Assim, aproveitou-se de ferramentas WEB, como: Google Scholar, onde foram levantados os autores e os conteúdos; AWS (Amazon Web Services) para disponibilização dos dados em um serviço online; Facebook Messenger para disponibilização do Aerobot, chatbot sobre ocorrências aéreas; além das ferramentas de nuvem usadas para desenvolver o artigo e para armazenar as fontes de dados. Todo esse ambiente online foi usado para o desenvolvimento desta pesquisa e da aplicação.

### **3.3 OBJETO DA PESQUISA**

Como já foi possível perceber, os objetos de estudo do presente artigo são os Dados Governamentais Abertos, buscando trazer novos conhecimentos sobre esse tema para os leitores. Com esse maior interesse, os cidadãos terão a possibilidade de fiscalizar temas de interesse da população, acarretando em uma evolução socioeconômica da sociedade brasileira. Desse modo, consegue-se fomentar o desenvolvimento de novos estudos sobre esse tema e novas aplicações podem surgir.

### 3.4 DELIMITAÇÃO E UNIVERSO DA AMOSTRA

Para a aplicação desenvolvida, foi necessário concentrar-se em uma base de dados específica. Para isso, escolheu-se três bases. Dentre elas, analisou-se e descobriu-se qual era de fato classificada como Dados Abertos. A partir disso, dentre todas as bases armazenadas online presentes no *dados.gov.br*, escolheu-se a sobre ocorrências aéreas no Brasil, preenchida e cuidada pelo **Cenipa**, Centro de Investigação e Prevenção de Acidentes Aéreos. Dentre elas, a aplicação preocupou-se com ocorrências que acometeram a aviação comercial, aviões comerciais. Desse modo, não se preocupou no momento em exibir informações sobre helicópteros, jatos particulares, entre outros meios de transporte aéreo. Outra razão para limitar quais dados seriam utilizados é justamente a limitação do servidor de dados, uma vez que, por ser gratuito, não aceita grande quantidade de dados.

### 3.5 INSTRUMENTO DE COLETA DE DADOS

Como dito anteriormente, nesta pesquisa científica não houve intervenção humana, não sendo necessário deslocar-se, conversar com pessoas na rua. Logo, não se aplicou formulários, questionários, entrevistas. Os instrumentos para coletar dados foram o Google Scholar e o repositório *dados.gov.br*, sendo o primeiro responsável pelo estudo bibliográfico e o segundo, pelo levantamento dos dados sobre aviação civil. É válido dizer que, tanto a análise quanto o armazenamento deles são pertinentes ao órgão público Cenipa. Esta pesquisa apenas utiliza os dados deste órgão, não se preocupando em que formato estão armazenados (importando apenas se é interpretável por máquina) e quais são os responsáveis por isso.

### 3.6 PROCEDIMENTOS METODOLÓGICOS

Para o desenvolvimento desse artigo, foi necessário primeiramente realizar o levantamento bibliográfico, buscando autores que abordassem temas correlatos ao conteúdo desta pesquisa, como: Dados Abertos, Ciência de Dados, Big Data, entre outros. Assim, buscou-se apresentar conceitos utilizados, também, para analisar bases de Dados Abertos Governamentais do *dados.gov.br*. Toda essa bibliografia foi levantada a partir de artigos disponíveis na WEB.

Diante da análise feita em cima das três bases de dados abertos neste artigo, foi escolhida a sobre ocorrências aéreas, disponibilizada pelo Cenipa. Ela foi utilizada para que o Aerobot - Chatbot de Aviação Civil - pudesse conversar com os clientes, usuários que entram em contato com o robô. Sendo assim, a máquina poderia apresentar informações aos seus clientes e tirar dúvidas sobre a aviação comercial e sobre sua segurança. Com essa aplicação, é possível trazer Dados Abertos, que muitas vezes não chegam aos ouvidos das pessoas, para a população. Para isso funcionar, foi necessário criar uma página do Facebook - **Aerobot** - para que as pessoas pudessem interagir com o robô (Figura 1 no Apêndice A). Também, a partir da configuração dela, algumas variáveis de conexão com a aplicação foram armazenadas para futuro uso.

Com relação à modelagem do Banco de Dados, utilizou-se a disponibilizada pelo próprio Cenipa (Figura 2 no Apêndice B), de modo que pudesse estar mais de acordo possível. Entretanto, na hora de transformá-la em código, foi necessário realizar algumas alterações nessa modelagem, uma vez que havia dados repetidos, o que acarretaria em mais espaço de memória com dados desnecessários. A modelagem foi um passo importante para compreender como os dados estavam estruturados e quais eram os seus relacionamentos. Assim, foi possível entender exatamente o fluxo das informações.

Após a modelagem e a definição do problema que a aplicação devia resolver, buscou-se a maneira como o software deveria funcionar. O primeiro passo foi desenvolver uma **API** - Application Programming Interface - responsável por receber as requisições dos clientes e retornar a informação desejada (Figura 3 no Apêndice C). Desse modo, o cliente, sabendo dos códigos das ocorrências aéreas, pode digitar um para a aplicação. Desse modo, a API vai verificá-lo e retornar as informações pertinentes à ocorrência selecionada. Para realizar esse processo, a API utiliza métodos GET para buscar os dados em um banco de dados armazenado na nuvem. Para o desenvolvimento dessa API e para criar o acesso ao banco online, utilizou-se a ferramenta Sails Js.

Com a API pronta, foi necessário entender como colocar o chatbot para comunicar-se com a interface. Desse modo, estudou-se o projeto disponibilizado pelo Facebook Messenger (Figura 4 no Apêndice D) como base para compreender o fluxo de como o dado chega, como ele é tratado e apresentado ao cliente final, usuário que entrou em contato com o Aerobot. Foram realizadas alterações neste projeto base, de modo que, quando o robô recebesse o código da ocorrência, ele requisitasse o banco de dados online. Assim, ele poderia recebê-lo e apresentá-lo ao cliente no messenger do Facebook.

Até o momento, todos os softwares gerados eram testados localmente, na própria máquina. Desse modo, pessoas com máquinas de outro IP não poderiam utilizar o serviço. Logo, foi necessário subir os programas em um servidor online gratuito, mesmo conhecendo suas limitações de memória, processamento, acesso. Também, foi necessário subir o banco de dados em um serviço online. Para isso, utilizou-se o **AWS**, Amazon Web Services (Figura 5 no Apêndice E). É válido dizer que se utilizou um plano gratuito da Amazon, o que também gera um *gargalo* - dificuldade - quanto à memória. É importante dizer que o tempo de resposta, no início das conversas e quando se passa um tempo sem falar com o robô, é um pouco maior, o que faz parte do pacote gratuito. Mas, para fins de estudo, essa gratuidade não é um problema.

Após todo esse processo de desenvolvimento da solução, realizou-se testes com alguns códigos de ocorrência existentes para entender o que a solução retornava. Verificou-se que estava trazendo as informações corretas - íntegra - para os clientes finais da solução. Caso o usuário digite um código errado, o chatbot exibe uma mensagem de erro e pergunta se ele deseja digitar novamente ou finalizar a consulta. Dessa forma, consegue-se ter interação com os clientes, como se de fato eles estivessem conversando com um ser humano do outro lado da tela. Assim, pode-se dizer que a solução atingiu seu propósito de distribuir os Dados Abertos do Cenipa, gerando um software útil para tirar dúvidas das pessoas sobre aviação civil (Figuras 6, 7 e 8 no Apêndice F).

#### **4. RESULTADOS E DISCUSSÃO**

A partir da pesquisa realizada, foi possível construir o chatbot (Aerobot) de forma gratuita. Para isso, escolheu-se ferramentas e aplicação de conversa que possibilitava a disponibilização dos produtos gerados. Para fazer a interação com nossos clientes, pessoas que usam a solução, utilizou-se o Facebook Messenger. Não foi possível todavia liberar o acesso do Aerobot para todas as pessoas. Isso ocorre, uma vez que a empresa Facebook solicita a análise da aplicação para verificar se ela está de acordo com sua política de privacidade. Logo, a aplicação se encontra online, mas ainda só responde aos usuários desenvolvedores da solução.

Foi percebido que a base de dados de ocorrências aéreas utilizada de fato contém dados abertos. Pode-se concluir isso, porque os dados presentes nela estão de acordo com todos os princípios levantados pelos 30 defensores da Califórnia, conforme verificado na pesquisa bibliográfica. Apesar de terem sido retirados os códigos de ocorrência repetidos da base, conseguiu-se corroborar com o que foi dito por Córdova, 2013 - o qual falou que a organização de dados em um repositório facilita a leitura, organização e disseminação deles.

## **5. CONSIDERAÇÕES FINAIS**

Conclui-se que foi desafiante trabalhar com ferramentas gratuitas, por conta das limitações de armazenamento, acesso, tempo de resposta existentes. No futuro, para disponibilização em massa, deve-se utilizar ferramentas pagas, que dão mais “poder” à solução. Pode-se, também, em outras versões disponibilizar a solução em outras aplicações de mensagens, como Whatsapp e Telegram. Para isso, deve-se solicitar a análise da solução, de modo que se possa verificar se o chatbot atende a todos os requisitos necessários. Por último, mas não menos importante, percebeu-se o quão é importante utilizar soluções para disseminar informação para as pessoas, de modo que elas possam fiscalizar dados obtidos em diversas mídias.



## REFERÊNCIAS

ANP. ANP - Série histórica de preços de combustíveis - revenda. **Portal Brasileiro de Dados Abertos**, 2021. Disponível em:  
<https://dados.gov.br/dataset/serie-historica-de-precos-de-combustiveis-por-revenda>. Acesso em: 12 ago. 2021

BRASIL. Decreto no. 7.724, de 16 de maio de 2012. Regulamenta a Lei no. 17.527, de 18 de novembro de 2011. Disponível em  
[http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2012/Decreto/D7724.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/Decreto/D7724.htm). Acesso em: 20 out. 2020.

BRASIL. Decreto no. 8.777, de 11 de maio de 2016. Institui a Política de Dados Abertos do Poder Executivo federal. Disponível em  
[http://www.planalto.gov.br/CCIVIL\\_03/\\_Ato2015-2018/2016/Decreto/D8777.htm](http://www.planalto.gov.br/CCIVIL_03/_Ato2015-2018/2016/Decreto/D8777.htm). Acesso em: 20 out. 2020.

BRASIL. Lei no. 17.527, de 18 de novembro de 2011. Lei de Acesso à Informação. Disponível em [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/l12527.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm). Acesso em: 20 out. 2020.

CAVIQUE, Luís. Big data e data science. Boletim da APDIO, p. 11-14, 2014.

CENIPA. CENIPA - Ocorrências Aeronáuticas na Aviação Civil Brasileira. **Portal Brasileiro de Dados Abertos**, 2021. Disponível em:  
<https://dados.gov.br/dataset/ocorrencias-aeronauticas-da-aviacao-civil-brasileira>. Acesso em: 12 ago. 2021

DE ALENCAR OLIVEIRA, Lairson E. Rodrigues; LÓSCIO, Bernadette Farias. Uma abordagem para captura de informações sobre aplicações que fazem uso de dados abertos. **Revista Brasileira de Administração Científica**, v. 5, n. 2, p. 127-140, 2014.

DOS PASSOS, Danielle Sandler. Big Data, Data Science e seus contributos para o avanço no uso da Open Source Intelligence. **Sistemas & Gestão**, v. 11, n. 4, p. 392-396, 2016.

DUTRA, Cláudio Crossetti; LOPES, Karen Maria Gross. Dados abertos: Uma forma inovadora de transparência. 2013

EZABELLA, Fernanda. País mais digital do mundo, Estônia 'hackeia' soluções para pandemia. ECOA, 2020. Disponível em:  
<https://www.uol.com.br/ecoa/ultimas-noticias/2020/04/11/como-o-pais-mais-digital-do-mundo-usa-a-inovacao-na-crise-do-coronavirus.htm>. Acesso em: 20 out. 2020

KLEIN, Rodrigo Hickmann; LUCIANO, Edimara Mezzomo; MACADAR, Marie Anne. Grau de transparência de dados abertos governamentais do site dados. rs. gov. br. **Revista Economia & Gestão**, 2015.

MINISTÉRIO DA SAÚDE. Coronavírus (covid-19). **Portal Brasileiro de Dados Abertos**, 2020. Disponível em: [https://dados.gov.br/dataset/notificacao\\_covid](https://dados.gov.br/dataset/notificacao_covid). Acesso em: 12 ago. 2021

PORTAL BRASILEIRO DE DADOS ABERTOS. Disponível em: <http://dados.gov.br/>. Acesso em: 20 out. 2020

PORTAL DE DADOS ABERTOS DA ESTÔNIA. Disponível em: <https://opendata.riik.ee/en/>. Acesso em: 20 out. 2020

PORTAL DE DADOS ABERTOS DA UNIÃO EUROPEIA. Disponível em:  
<http://data.europa.eu/euodp/pt>. Acesso em: 20 out. 2020

PORTAL DE DADOS ABERTOS DO REINO UNIDO. Disponível em:<https://data.gov.uk/>. Acesso em: 20 out. 2020

PORTAL DE DADOS ABERTOS DOS ESTADOS UNIDOS. Disponível em: <https://data.gov/>.  
Acesso em: 20 out. 2020

POSSAMAI, Ana Júlia; DE SOUZA, Vitoria Gonzatti. Transparência e Dados Abertos Governamentais: Possibilidades e Desafios a Partir da Lei De Acesso À Informação. **Administração Pública e Gestão Social**, 2020.

RAUTENBERG, Sandro; DO CARMO, Paulo Ricardo Viviurka. Big data e ciência de dados. **Brazilian Journal of Information Science: research trends**, v. 13, n. 1, p. 56-67, 2019.

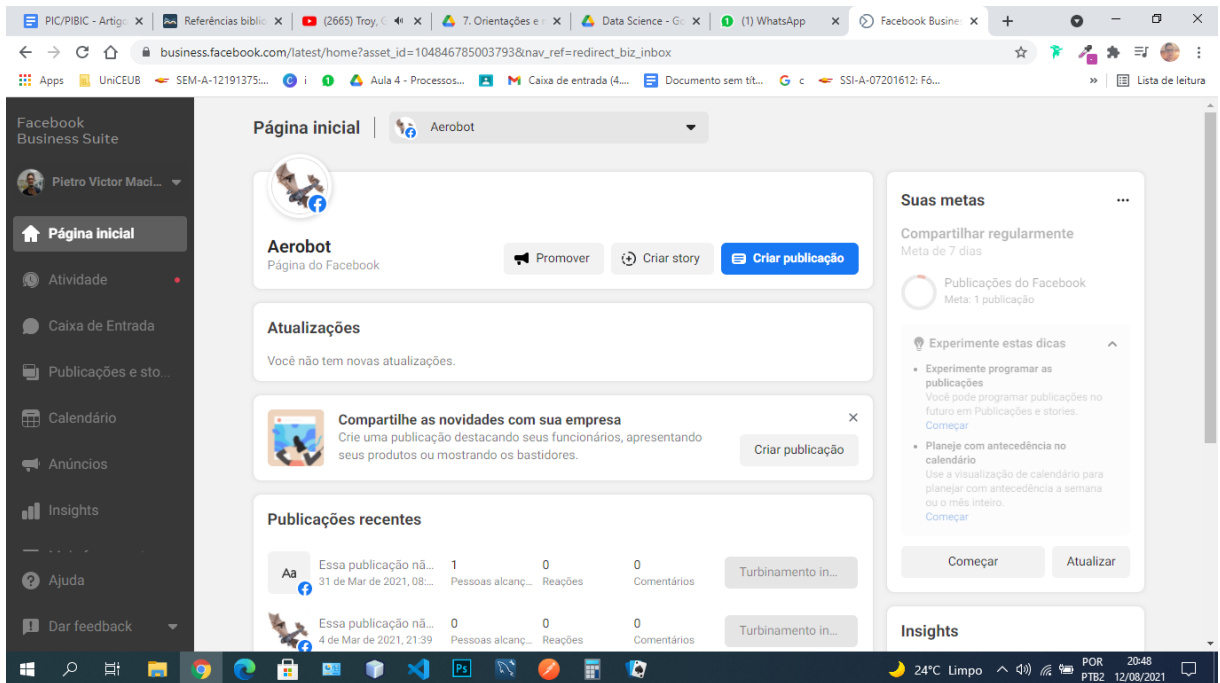
RONDON, Thiago Berlitz; KOGAN, Ariel. Democracia em nuvens: A e-Estônia inspira participação cidadã por meio da tecnologia. Politize!, 2017. Disponível em: <https://www.politize.com.br/e-estonia-democracia-em-nuvens/>. Acesso em: 20 out. 2020

TELES MOURA, Aristóteles Lamartine; GOMES AMORIM, Dinani. BIG DATA: O IMPACTO E SUA FUNCIONALIDADE NA SOCIEDADE TECNOLÓGICA. **Revista OPARA**, v. 4, n. 1, 2014.

VAZ, José Carlos; RIBEIRO, Manuella Maia; MATHEUS, Ricardo. Dados governamentais abertos e seus impactos sobre os conceitos e práticas de transparência no Brasil. **Cadernos ppg-au/ufba**, v. 9, n. 1, 2010.

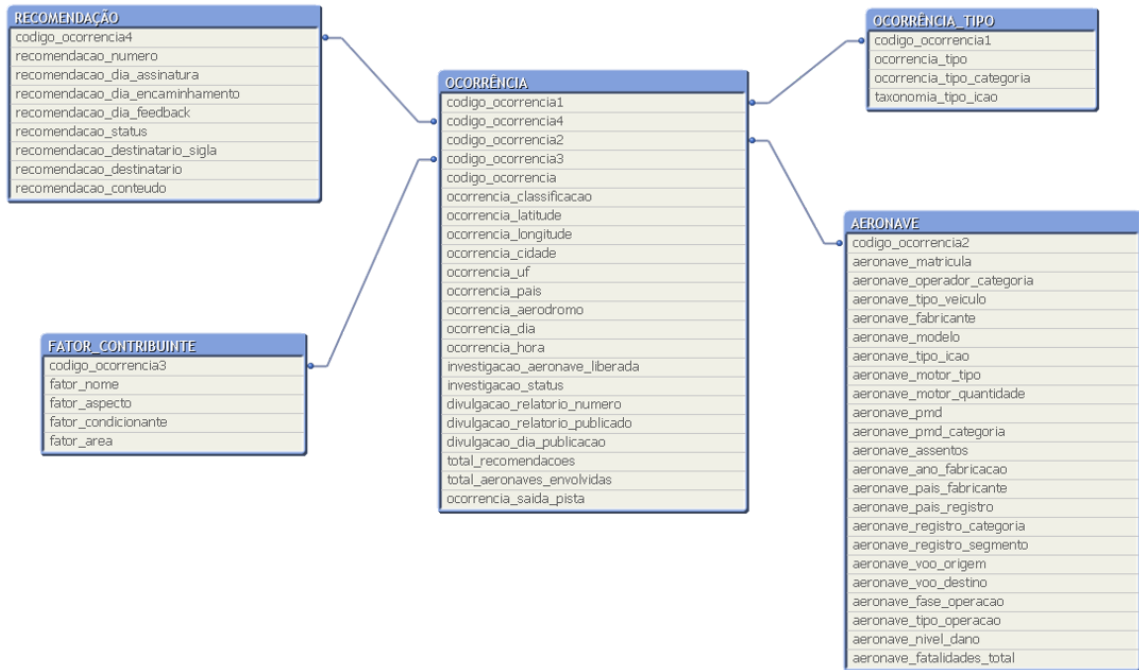
# APÊNDICE A - PÁGINA DO AEROBOT

Figura 1.



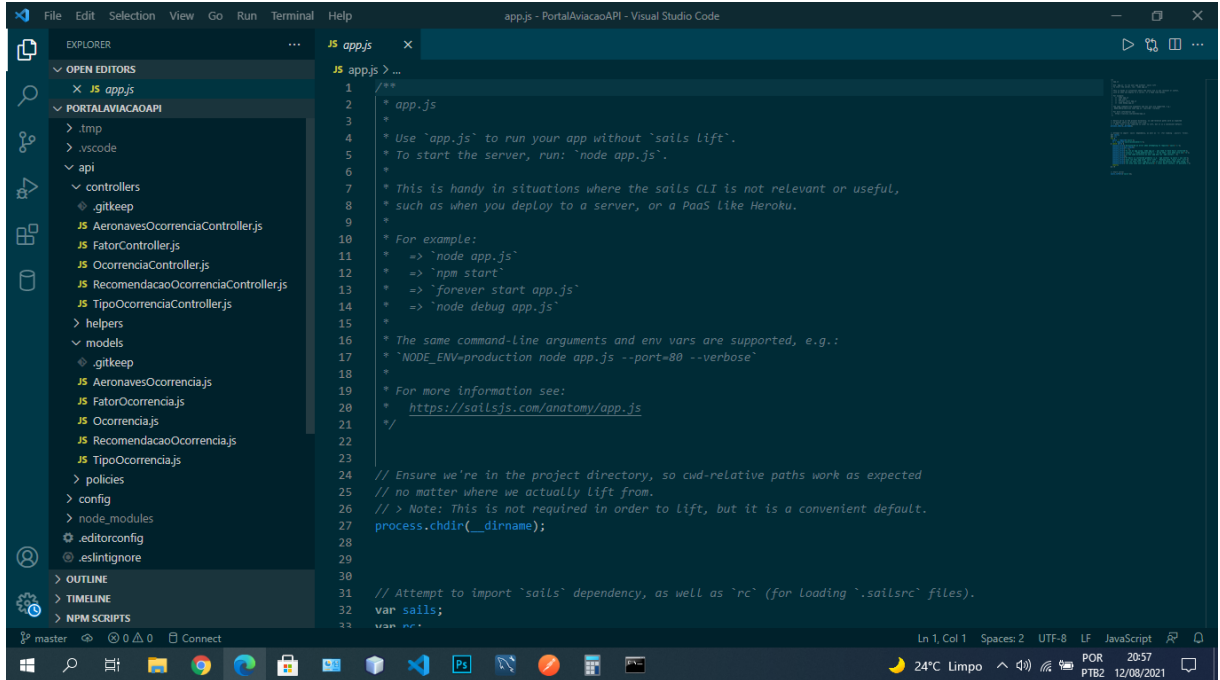
## APÊNDICE B - MODELAGEM DO BANCO DE DADOS DO CENIPA

Figura 2.



## APÊNDICE C - API DE AVIAÇÃO COMERCIAL DESENVOLVIDA

Figura 3.



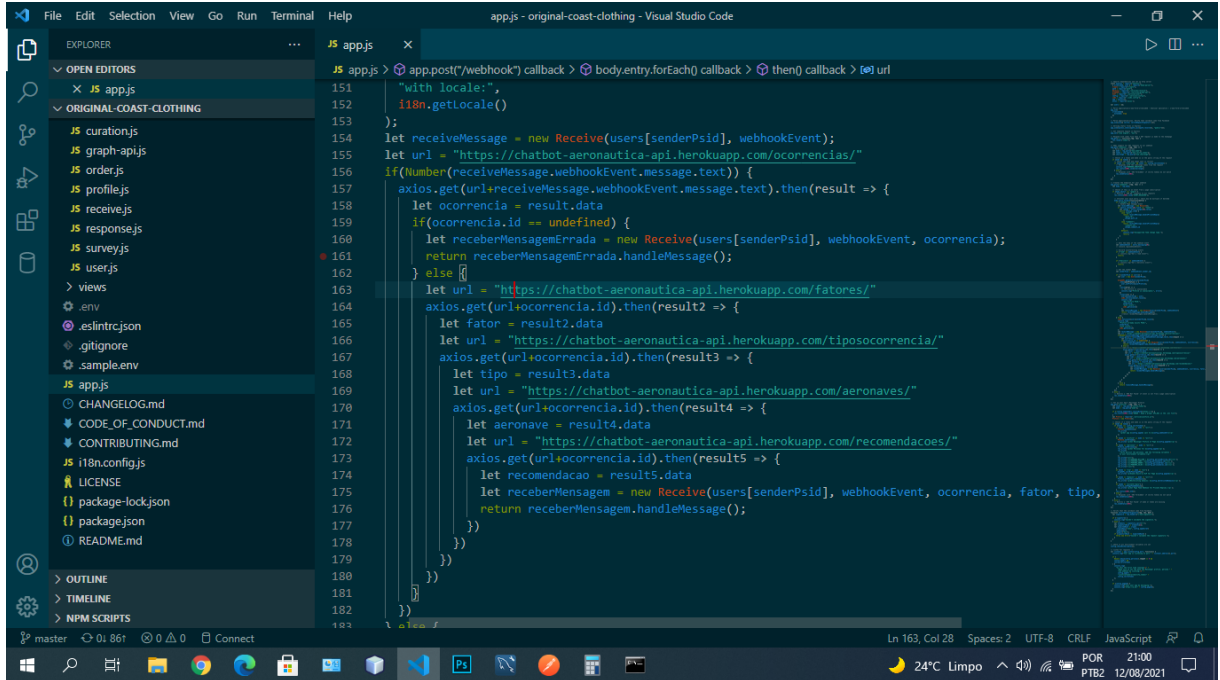
The image shows a screenshot of the Visual Studio Code editor interface. The Explorer sidebar on the left displays the project structure for 'PORTALAVIACA-API', including folders like 'api', 'controllers', 'models', and 'policies', and various JavaScript files. The main editor window is open to the 'app.js' file, which contains the following code:

```
1  /**
2  * app.js
3  *
4  * Use `app.js` to run your app without `sails lift`.
5  * To start the server, run: `node app.js`.
6  *
7  * This is handy in situations where the sails CLI is not relevant or useful,
8  * such as when you deploy to a server, or a PaaS like Heroku.
9  *
10 * For example:
11 * => `node app.js`
12 * => `npm start`
13 * => `forever start app.js`
14 * => `node debug app.js`
15 *
16 * The same command-line arguments and env vars are supported, e.g.:
17 * `NODE_ENV=production node app.js --port=80 --verbose`
18 *
19 * For more information see:
20 * https://sailsjs.com/anatomy/app.js
21 */
22
23 // Ensure we're in the project directory, so cwd-relative paths work as expected
24 // no matter where we actually lift from.
25 // > Note: This is not required in order to lift, but it is a convenient default.
26 process.chdir(__dirname);
27
28
29
30
31 // Attempt to import `sails` dependency, as well as `rc` (for loading `.sailsrc` files).
32 var sails;
33
```

The status bar at the bottom indicates the current file is 'app.js', line 1, column 1, with 2 spaces, UTF-8 encoding, and JavaScript language mode. The system tray shows the date and time as 20:57 on 12/08/2021.

## APÊNDICE D - CODIFICAÇÃO DO CHATBOT

Figura 4.



```
151     "with locale:",
152     i18n.getLocale()
153   });
154   let receiveMessage = new Receive(users[senderPsid], webhookEvent);
155   let url = "https://chatbot-aeronautica-api.herokuapp.com/ocorrencias/"
156   if(Number(receiveMessage.webhookEvent.message.text)) {
157     axios.get(url+receiveMessage.webhookEvent.message.text).then(result => {
158       let ocorrencia = result.data
159       if(ocorrencia.id == undefined) {
160         let receberMensagemErrada = new Receive(users[senderPsid], webhookEvent, ocorrencia);
161         return receberMensagemErrada.handleMessage();
162       } else {
163         let url = "https://chatbot-aeronautica-api.herokuapp.com/fatores/"
164         axios.get(url+ocorrencia.id).then(result2 => {
165           let fator = result2.data
166           let url = "https://chatbot-aeronautica-api.herokuapp.com/tiposocorrencia/"
167           axios.get(url+ocorrencia.id).then(result3 => {
168             let tipo = result3.data
169             let url = "https://chatbot-aeronautica-api.herokuapp.com/aeronaves/"
170             axios.get(url+ocorrencia.id).then(result4 => {
171               let aeronave = result4.data
172               let url = "https://chatbot-aeronautica-api.herokuapp.com/recomendacoes/"
173               axios.get(url+ocorrencia.id).then(results => {
174                 let recomendacao = results.data
175                 let receberMensagem = new Receive(users[senderPsid], webhookEvent, ocorrencia, fator, tipo,
176                 recomendacao);
177                 return receberMensagem.handleMessage();
178               })
179             })
180           })
181         })
182       })
183     })
  }
```

## APÊNDICE E - SERVIÇO AWS PARA HOSPEDAGEM DOS DADOS

Figura 5.

The screenshot shows the AWS Management Console interface for Amazon RDS. The left sidebar contains navigation options such as Dashboard, Databases, Query Editor, Performance Insights, Snapshots, Automated backups, Reserved instances, Proxies, Subnet groups, Parameter groups, Option groups, Events, Event subscriptions, and Recommendations. The main content area is titled 'RDS > Databases' and features a 'Group resources' toggle, 'Modify', 'Actions', 'Restore from S3', and 'Create database' buttons. A search bar labeled 'Filter databases' is present. Below is a table listing database instances:

DB identifier	Role	Engine	Region & AZ	Size	Status
mysqlserver	Instance	MySQL Community	us-east-2b	db.t2.micro	Available

The bottom of the image shows the Windows taskbar with various application icons and system tray information including temperature (24°C), location (Limpo), and date/time (12/08/2021).



## APÊNDICE F - TESTES NO AEROBOT

Figura 6.

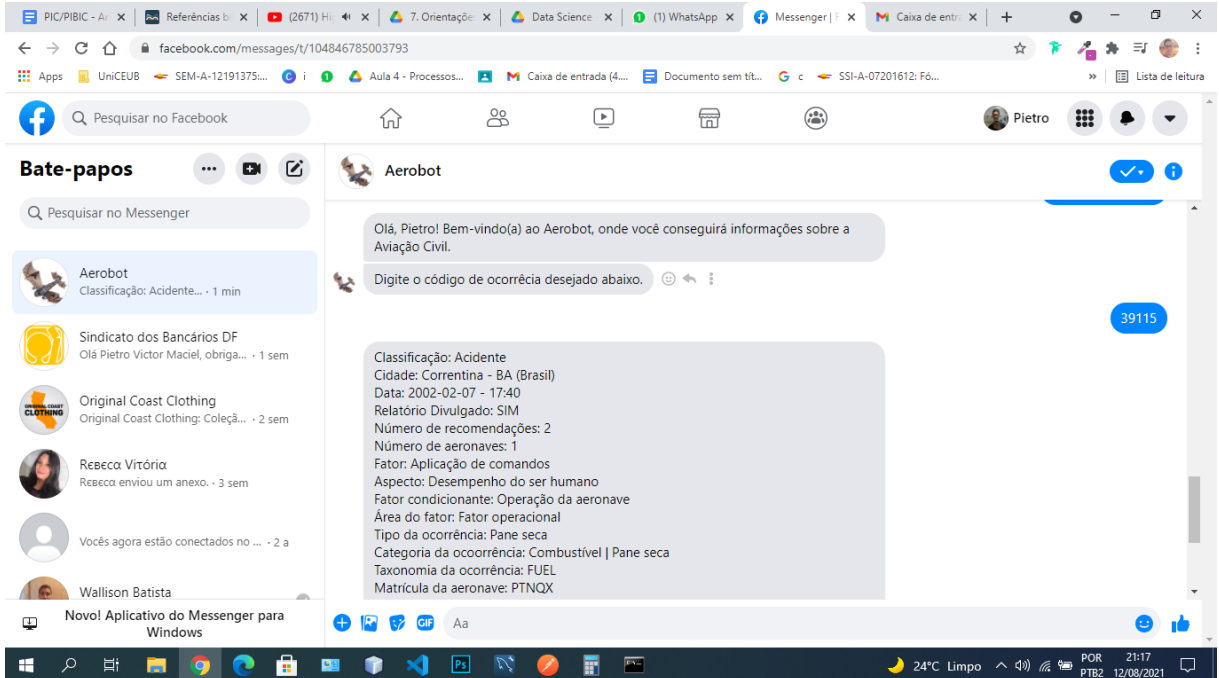


Figura 7.

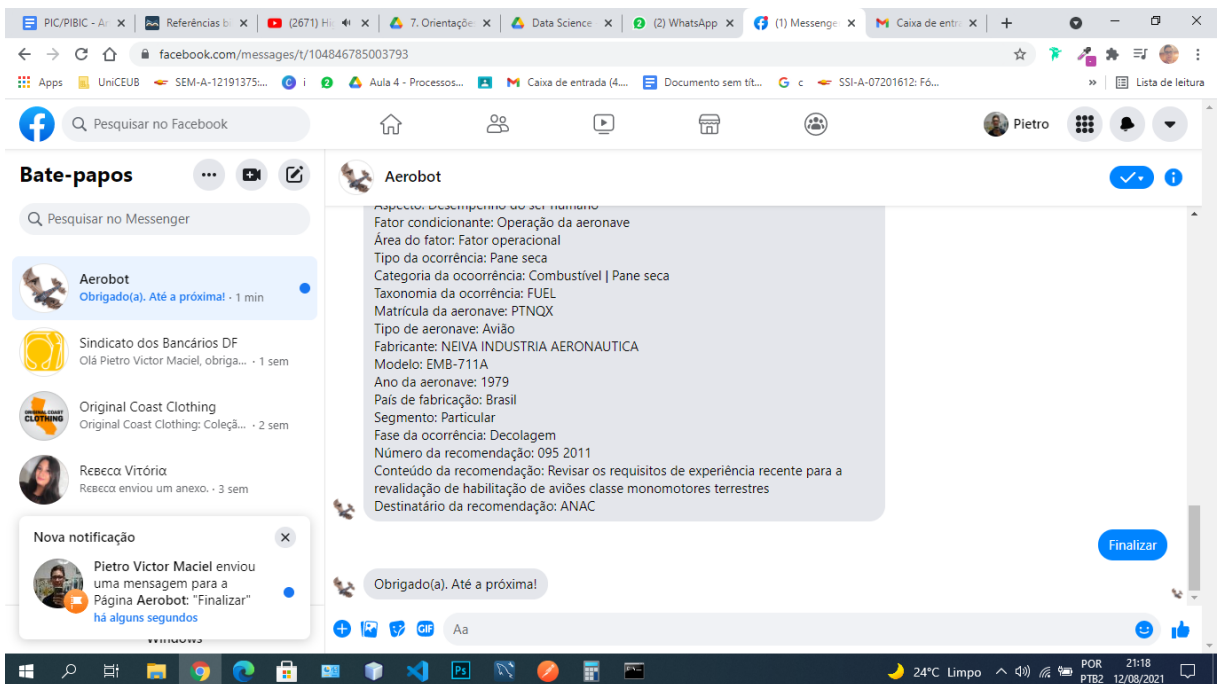


Figura 8.

