

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE
SÃO PAULO CÂMPUS CAMPINAS

GIOVANNI FONSECA CORREIA

**IDENTIFICAÇÃO E EXTRAÇÃO DE DADOS DE PORTAIS DE DADOS
ABERTOS: UMA INICIATIVA *OPEN DATA***

CAMPINAS

2019

GIOVANNI FONSECA CORREIA

**IDENTIFICAÇÃO E EXTRAÇÃO DE DADOS DE PORTAIS DE DADOS
ABERTOS: UMA INICIATIVA *OPEN DATA***

Trabalho de Conclusão de Curso apresentado como exigência parcial para obtenção do diploma do Curso de Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia Câmpus Campinas.

Orientador: Prof. Dr. Andreiuid Sheffer Corrêa

CAMPINAS

2019

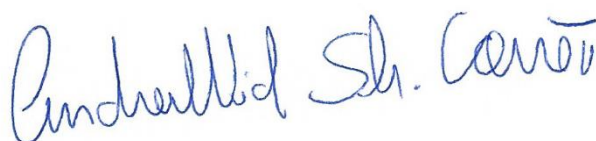
Giovanni Fonseca Correia

**Identificação e extração de dados de portais de dados abertos: uma
iniciativa *Open Data***

Trabalho de Conclusão de Curso apresentado como exigência parcial para obtenção do diploma do Curso de Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo Câmpus Campinas.

Aprovado pela banca examinadora em: 07 de novembro de 2019.

BANCA EXAMINADORA



Prof. Dr. Andreiwid Sheffer Corrêa (orientador)

IFSP Câmpus Campinas

Prof. Dr Ricardo Barz Sovat

IFSP Câmpus Campinas

Prof. Me Everton Josué da Silva

IFSP Câmpus Campinas

*Dedico este trabalho aos meus familiares,
colegas de classe, professores e servidores do Instituto
que colaboraram em minha jornada formativa.*

Ficha Catalográfica
Instituto Federal de São Paulo – Câmpus Campinas
Biblioteca
Danielle Sarmento – CRB 8/8669

C824i Correia, Giovanni Fonseca
 Identificação e extração de dados de portais de dados abertos: uma iniciativa
 Open Data. / Giovanni Fonseca Correia. – Campinas, SP: [s.n.], 2019.
 51f. : il.

Orientador: Andreiuid Sheffer Corrêa.
Trabalho de Conclusão de Curso (graduação) – Instituto Federal de
Educação, Ciência e Tecnologia de São Paulo Câmpus Campinas. Curso de
Tecnologia em Análise e Desenvolvimento de Sistemas, 2019.

1. Dados abertos. 2. Plataformas. 3. Algoritmos. 4. Python. I. Instituto Federal
de Educação, Ciência e Tecnologia de São Paulo Câmpus Campinas. Curso de
Tecnologia em Análise e Desenvolvimento de Sistemas. II. Título.

AGRADECIMENTOS

Agradeço primeiramente a Deus, pelo dom da vida e pela oportunidade de concluir mais uma etapa de minha experiência acadêmica.

Agradeço a todos os professores e servidores do IFSP
Câmpus Campinas, que contribuíram direta e
indiretamente para a conclusão desse trabalho.

Agradeço também à minha família, que deu todo o apoio
necessário para que eu chegasse até aqui.

Agradeço ao meu orientador que me auxiliou a solucionar as
dificuldades encontradas no caminho.

RESUMO

Com a crescente do tema de dados abertos nos últimos anos e com a disponibilização de cada vez mais portais que buscam atender aos requisitos de abertura de dados, verificou-se a necessidade de um método que seja capaz de identificar qual a plataforma utilizada para o desenvolvimento dos portais de dados, e se suas ações condizem com a política dos dados abertos. Assim, o projeto tem como principal objetivo o desenvolvimento de algoritmos capazes de identificar qual a plataforma utilizada e extrair informações básicas como a quantidade de *datasets* e a localização da iniciativa. Uma vez selecionadas as plataformas a serem consideradas, este projeto contribui com uma comparação entre elas, especificando os principais recursos disponibilizados, de modo que a melhor solução seja indicada a partir de um conjunto de necessidades.

Palavras-chave: Dados abertos. Plataformas. Algoritmos. Python.

ABSTRACT

With the growing open data theme in recent years and the availability of more portals that seek to meet the requirements of data opening, there was a need for a method that is able to identify which platform is used for development portals, and whether their actions match the open data policy. Thus, the project has as its main objective the development of algorithms capable of identifying which platform is used and extracting basic information such as the number of datasets and the location of the initiative. Once the platforms to be considered were selected, this project contributes a comparison between them, specifying the main resources available, so that the best solution is indicated from a set of needs.

Keywords: Open data. Platforms. Algorithms. Python.

LISTA DE FIGURAS

Figura 1 – Metodologia implantada.....	23
Figura 2 – Quantidade de portais identificados.....	39
Figura 3 – Quantidade de portais identificados por plataforma.....	40
Figura 4 – Quantidade de portais identificados.....	41
Figura 5 – Quantidade de portais identificados por plataforma.....	42
Figura 6 – Média de <i>datasets</i> por plataforma identificada.....	43
Figura 7 – Média de <i>datasets</i> por plataforma identificada.....	44

LISTA DE TABELAS

Tabela 1 – Comparação entre plataformas de dados abertos.....	24
---	----

SUMÁRIO

1 INTRODUÇÃO.....	12
2 CONTEÚDO DA PARTE TEXTUAL.....	13
2.1 JUSTIFICATIVA.....	13
2.2 OBJETIVOS.....	14
2.3 FUNDAMENTAÇÃO TEÓRICA.....	15
3 MÉTODO.....	23
4 RESULTADOS.....	39
5 CONSIDERAÇÕES FINAIS.....	45
6 REFERÊNCIAS.....	47
APÊNDICE A - Tabela comparativa entre as plataformas acerca de categorização e apresentação visual.....	49

1 INTRODUÇÃO

O tema dados abertos vem crescendo bastante nos últimos tempos, apesar de não alcançar, em alguns países, o patamar esperado para viabilizar a disponibilização irrestrita de dados. A pressão de movimentos e da própria sociedade pela disponibilização dos dados de seus governos de forma pública fez com que a utilização de plataformas de dados abertos crescesse e fosse a principal motivação para a realização dessa monografia.

Há uma série de trabalhos realizados na área de identificação de portais de dados abertos, porém, foram encontrados apenas maneiras manuais de identificar portais governamentais utilizando plataformas de dados abertos. Há um projeto, por exemplo, chamado Dataopen.eu que fornece em forma de texto uma pesquisa detalhada especificamente para portais governamentais, tendo como promessa um recurso para descobrir conteúdo oculto em seus sites. Porém o trabalho não fornece muitos detalhes sobre a implementação do código, tampouco como obter dados acerca dos portais. Assim, este trabalho tem como objetivo principal desenvolver algoritmos capazes de identificar e checar essas plataformas governamentais de maneira automática, descobrindo qual o software utilizado para implementação do portal, além de extrair informações sobre a quantidade de *datasets* disponibilizadas por aquele portal.

Como principal resultado, tem-se a comprovação da eficácia dos algoritmos em um teste de massas de URL, apontando as plataformas de dados abertos de cada um, caso tenha sido utilizada alguma, contribuindo assim com a automatização da verificação de plataformas de portais de dados abertos.

O trabalho encontra-se organizado em 6 Sessões: Introdução, Conteúdo da parte textual, essa dividida em Justificativa, Objetivos e Fundamentação teórica, Metodologia, Resultados, Considerações finais e Referências.

2 CONTEÚDO DA PARTE TEXTUAL

2.1 JUSTIFICATIVA

O trabalho tem como motivação principal o fato de não terem sido encontrados trabalhos que realizam de forma automática a identificação de portais que utilizam plataformas de dados abertos. Verifica-se projetos mundiais que são desenvolvidos na área, mas que envolvem um custo grande em função de sua projeção global, seja de recursos físicos ou computacionais. O *Open Data Barometer*, por exemplo, em sua 4ª edição, traz uma noção global de como os governos estão publicando e usando seus dados abertos e analisa membros do G20 que se comprometeram com os princípios de dados abertos anticorrupção do G20. Também pode-se citar o *Global Open Data Index* como outro trabalho nessa área, que realiza uma análise acerca dos dados abertos governamentais e disponibiliza um ranking para a comparação entre os países.

Este trabalho não vem para substituir esses projetos globais nem oferecer os mesmos recursos, porém apresenta-se como uma alternativa automatizada para se realizar o processo de identificação e extração de dados dos portais governamentais, visto que processos manuais podem se tornar muito custosos em função do tempo gasto na identificação de portal por portal. Além disso, podem tornar-se obsoletos com o tempo, pois portais podem mudar a qualquer momento para outra plataforma de dados abertos e, como dito anteriormente, uma nova identificação manual seria muito custosa.

2.2 OBJETIVOS

2.2.1 OBJETIVO GERAL

Identificar portais de dados abertos a partir do desenvolvimento de assinaturas, verificação da disponibilidade e análise automatizada das plataformas de software mais utilizadas na área.

2.2.2 OBJETIVOS ESPECÍFICOS

- a) Levantar as principais plataformas de dados abertos em utilização;
- b) Mapear as características e meios de acesso;
- c) Desenvolver os algoritmos de identificação de plataformas;
- d) Extrair dados de utilização.

2.3 FUNDAMENTAÇÃO TEÓRICA

2.3.1 *Open Government Data*

Entende-se com dados abertos textos, imagens, fotos, planilhas, que estão disponíveis para qualquer pessoa, estão livres para serem utilizados, distribuídos, sujeitos, no máximo, a terem que referenciar a legítima fonte (OPEN..., 2012).

O *Open Government Data* (OGD) é um movimento que busca promover a transparência dos registros públicos, fiscalizá-los e, responsabilizá-los por erros e promessas não cumpridas. Além disso, com a disponibilização dos dados de forma aberta à população, o OGD espera que isso cause uma reflexão maior na sociedade, para que essa tome decisões mais embasadas e fundamentadas em informações verídicas.

Além do mais, o OGD também possui aspirações econômicas. De acordo com o movimento, a abertura de dados governamentais à população incentivaria o empreendedorismo e inovação. Segundo eles, abriria a possibilidade de corporações, indivíduos, empresas privadas, governamentais criarem produtos e serviços utilizando dados públicos.

Alguns exemplos que o movimento OGD considera relevantes:

- Informações comerciais;
- Dados sobre negociações oficiais do governo;
- Registros, patentes, marcas registradas;
- Banco de dados de concursos públicos;
- Informações geográficas;
- Informações legais, que incluem decisões dos tribunais nacionais, estrangeiros, legislações, tratados;
- Dados sociais, dos mais variados tipos, economia, saúde, desemprego;
- Dados meteorológicos;
- Informações de transporte, sobre tráfego, trânsito, obras (UBALDI, 2013).

2.3.2 Portais de dados abertos

Em alguns momentos podem ocorrer confusões acerca da definição de portais de dados abertos e plataformas de dados abertos. Plataformas de dados abertos são, em sua maioria, *softwares* vendidos como serviço, disponibilizando toda uma infraestrutura para um governo, uma corporação, uma empresa que deseja disponibilizar seus dados abertos. E os portais de dados abertos são justamente esses que através da internet disponibilizam seus dados para o acesso de todos.

Uma das críticas acerca de plataformas de dados abertos é a de publicações em formato *Portable Document Format* (PDF) ou em extensões que necessitam do pacote Office da Microsoft, por exemplo, pois isso pode limitar o acesso desses dados à comunidade.

Seguindo essa linha de raciocínio, em um relatório apresentado pela *Government Working Group* (2007), existem oito critérios necessários para que os dados governamentais sejam reconhecidos como dados abertos. Eles devem ser:

- Completos, ou seja, não possuem limitações de privacidade;
- Primários, disponíveis com o menor número possível de alterações;
- Atuais, estarem disponíveis assim que solicitados;
- Acessíveis, disponíveis para o maior número de pessoas e o maior número de finalidades possíveis;
- Processáveis por máquinas, ou seja, deve possibilitar o processamento automatizado;
- Não-discriminatórios, devem estar disponíveis para todos, sem exceção e nenhuma exigência prévia, por exemplo, cadastros, preenchimento de formulários;
- Não-proprietários, que estejam em um formato que não seja de posse de nenhuma entidade;
- Por fim, livres de licença, não haja nenhuma restrição ligada a patentes, direitos autorais.

A consolidação dos dados abertos poderia se tornar uma mudança significativa na relação governo-sociedade. Os portais de dados abertos seriam usados para uma interação com o intuito de uma colaboração mútua entre a população, empresas, corporações e os governantes. Porém, poucos governos utilizam a plataforma de dados abertos (CUNHA, 2015; ROSINA, 2015; TEIXEIRA, 2015; SILVA, 2015; LAZZARI, 2015; SILVA, 2015; KAROLCZAK, 2015; CANTONI, 2015; COELHO, 2015; SILVA, 2015; SPINOLA, 2015; MARINHO, 2015; SCHELIGA, 2015).

Segundo dados da *Global Report* (2017), após serem analisados 1.725 portais de 115 países diferentes, somente 7% de todos os dados eram totalmente abertos, um número muito aquém do esperado.

Contudo, segundo *United States Of America* (2016), esse número inexpressivo de portais que disponibilizam seus dados de forma aberta é, de certa maneira, compreensível. Não é tão fácil simplesmente abrir seus dados, há questões jurídicas, políticas, preocupação com a proteção e segurança desses dados. O projeto *UN- DESA* (2014) mapeou nove fatores necessários para uma implantação bem-sucedida de dados abertos. São eles:

- Compromisso do governo;
- *Frameworks* legais;
- Estruturas institucionais;
- Responsabilidades e capacidades dentro do governo;
- Políticas e procedimentos de gerenciamento de dados do governo;
- Demanda por dados abertos;
- Engajamento cívico e capacidades para dados abertos;
- Financiamento de um programa de dados aberto;
- Tecnologia nacional.

O Portal Brasileiro de Dados Abertos é o responsável por centralizar a busca a dados abertos no governo brasileiro. Desde o seu princípio, levou em conta a participação popular a um patamar bem alto, as reuniões sobre o planejamento e a execução do projeto de criação contavam com a participação de qualquer

interessado. Além disso, quando possível, as reuniões eram transmitidas ao vivo pela Internet. Feito para uso dos Três Poderes a nível Federal, Estadual, Distrital e Municipal, para os cidadãos e entidades privadas, sejam nacionais ou internacionais e inspirado por Estados Unidos e Reino Unido, os quais construíram seus portais de dados abertos em 2009 e 2010, respectivamente, o Portal Brasileiro de Dados Abertos, lançado em 2012, tinha como principais objetivos:

- Definir e padronizar os dados e suas técnicas de compartilhamento para uso do governo e da população.
- Buscar a evolução na publicação de dados abertos, seguindo práticas emergentes tanto no cenário nacional quanto no internacional.
- Promover a colaboração entre a sociedade e seus governantes
- Estimular a participação popular na construção de uma sociedade habituada com o reuso e agregação de valor dos dados públicos.

O orçamento limitado e a participação de colaboradores voluntários, dos quais não podia se exigir o mesmo empenho de uma equipe formada especificamente para este objetivo, foram os maiores desafios para a concretização do projeto. Esses obstáculos foram vencidos com a utilização de softwares livres e um escopo traçado de forma mais flexível. (BATISTA, 2015).

Vários aplicativos foram desenvolvidos por intermédio dos dados disponibilizados pelo Portal Brasileiro de Dados Abertos. Alguns deles citados no próprio site do Portal Brasileiro de Dados Abertos (2015).

- “Reputação S.A”, o qual traz informações acerca de várias empresas, com o intuito de ajudar na contratação de serviços, produtos.
- “Reclamações BR”, com um objetivo semelhante ao Reputação S.A, traz uma série de empresas e o número de reclamações em cada uma delas.
- “Para onde foi meu dinheiro”, um aplicativo que mostra graficamente para onde foi e como foi executado o orçamento do governo, na esfera federal e no estado de São Paulo.
- Um dos mais significativos, premiado na Maratona de Programação *Hackaton* da Câmara municipal de São Paulo, o aplicativo “Siga seu vereador”,

que traz uma linha do tempo com as ações, medidas de cada vereador do mandato atual. Cada usuário pode seguir os vereadores que desejar.

Os portais de dados abertos podem escolher para seu desenvolvimento uma plataforma de dados já existente no mercado ou, até mesmo, desenvolverem a sua própria, desde que atendam aos princípios já citados acima. Este trabalho focou nas plataformas já existentes no mercado e algumas delas estão descritas nas seções abaixo.

2.3.3 CKAN

O CKAN é uma ferramenta utilizada na construção de portais de dados abertos, que serve como apoio à publicação e gestão da coleta de dados. Desenvolvida por uma organização chamada *Open Knowledge Foundation*, é uma plataforma de código aberto constituída por uma comunidade de desenvolvedores, programadores que busca encontrar e corrigir erros. O CKAN tem como principais funções o compartilhamento e descoberta de dados. (CONEGLIAN, 2015; SEGUNDO, 2015).

Entre seus principais usuários, a fim de compartilhar suas informações e disponibilizá-las para o acesso de qualquer pessoa, estão governos nacionais, governos locais, casas legislativas e instituições de pesquisa. Entre os mais famosos internacionalmente estão o Portal de Dados Abertos do Reino Unido e dos EUA. Já a nível nacional, temos o portal do Senado Federal, do Rio Grande do Sul e Recife. O CKAN oferece versões em 53 idiomas. Em seu funcionamento, há uma série de funcionalidades interessantes ao usuário.

Após pesquisar todos os metadados e retornar com o resultado, o CKAN consegue catalogar dados similares ao procurado, como por exemplo, os mesmos dados, porém em formatos diferentes, ou períodos distintos.

Além disso, traz o resultado também em pré-visualizações em forma de gráficos, tabelas, permitindo ordenar a busca conforme a necessidade do usuário. O CKAN permite também que se realizem buscas em uma determinada área, por meio de um mapa ele permite que o usuário delimite a área onde os dados devem ser buscados. É possível consultar o histórico de atualizações, deixando claro quais alterações foram feitas no documento e o porquê, agrupar os dados em categorias entre outras funcionalidades. (HERRMANN, 2014).

2.3.4 Socrata

A ferramenta Socrata foi criada em 2007, por Kevin Merritt, e foi um dos softwares mais utilizados até o surgimento da CKAN, que passou a dominar o mercado. Atualmente, segundo dados da própria Socrata, cerca de 1.000 agências governamentais estão utilizando esta plataforma (CORRÊA, 2018; ZANDER, 2018; SILVA, 2018). E, conforme citado no artigo de Isotani (2018) e Bittencourt (2018), o Socrata foi desenvolvido com tecnologias mais recentes que muitas outras ferramentas da área, foi desenvolvido, por exemplo, utilizando MongoDB, ferramenta nova no mercado. Além disso, disponibiliza um conjunto de bibliotecas que permite seu desenvolvimento em mais de um tipo de linguagem, seja R, PHP, Java etc. E possui duas versões de uso, uma para o mercado privado e outra para a comunidade *Open Source*.

2.3.5 ArcGis Open Data

O ArcGis, criado pela ESRI (*Environmental Systems Research Institute*), é um conjunto de softwares integrados voltado para análise e processamento de dados geográficos. Constituído por outras 6 ferramentas, o ArcGis reúne e organiza informações conforme a semelhança de suas origens. E esses dados são representados de duas maneiras diferentes: No formato Vetorial ou no formato Raster. No formato Vetorial, os dados obtidos são representados por pontos, linhas ou polígonos. Já no formato Raster, as informações são guardadas em matrizes de tamanho igual. (SILVA, 2010; MACHADO, 2010).

Inicialmente, o produto ArcGis não foi desenvolvido como uma plataforma de dados abertos, como dito acima, o foco da ferramenta era a análise e o processamento de dados geográficos, mas com o passar dos anos uma funcionalidade se agregou a outra e temos alguns casos de sucesso da integração, alguns dos exemplos de uso do ArcGis nos EUA: O Departamento de Estado dos Serviços e de Saúde de Washington utiliza para configurar e expor dados acerca de saúde mental, casas de repouso etc. Dados que podem ser baixados ou visualizados de forma bem rápida por seus usuários. O Distrito de Columbia também utiliza o ArcGis para disponibilizar seus dados, e funciona a partir de sua localização no momento do uso. Outra instituição que faz uso do ArcGis é o

Departamento de Agricultura dos EUA (USDA), o qual faz o uso da plataforma para mapear áreas do país para combater a obesidade infantil. (ESRI, 2015).

2.3.6 Junar

Junar é mais uma das plataformas de dados abertos existentes no mercado. Em suma, grandes e pequenas organizações, indivíduos e até usuários fora de *firewall* podem navegar e fazer uso dos dados disponibilizados, seja apenas visualizando-os, ou baixando-os por completo. Entre as suas principais vantagens estão:

- Independência dos backends de aplicativos, suporte para todos os tipos de formato;
- Não requer recursos caros para instalação e uso;
- Dados compatíveis para serem baixados em qualquer dispositivo;
- API utilizada suporta tranquilamente o compartilhamento máquina a máquina. Agora com relação aos desenvolvedores, as principais funcionalidades oferecidas pela plataforma são:
 - Pesquisas nos catálogos de visualizações do site, seus gráficos e painéis;
 - Importação de dados via requisições HTTP GET;
 - Recuperação de exibições de dados e gráficos. E a API pode retornar dados nos seguintes formatos: HTML, CSV, XML, JSON etc.

Além disso, outro fator que torna atraente a plataforma Junar é o fato de o usuário ou a organização utilizadora não ter que se preocupar com armazenamento, hardwares, segurança etc. A Junar trabalha junto a Amazon Web Services (AWS), e garante 99,5% de disponibilidade e segurança de dados. (JUNAR, 2014).

2.3.7 APIs

Uma API é um conjunto de padrões programáveis com a intenção de comunicar softwares uns com os outros. Cada API funciona de uma maneira diferente, algumas contém padrões internacionais, outras contém bibliotecas específicas de linguagens de programação etc. (APPLICATION..., 2015)

Já pela definição de Pires (2013), uma API funciona como uma camada intermediária, que promove a interação entre uma base de dados e um aplicativo, uma página, um software que se alimenta desses dados. Uma das principais funções de uma API é possibilitar a extração de dados por meio de requisições Web, além disso, promove um acesso facilitado a dados específicos de uma base, não é necessário baixar todos os dados. As APIs podem ser classificadas em públicas ou privadas. Em APIs privadas, elas são criadas por desenvolvedores para facilitar o acesso aos dados desejados. Já as públicas são desenvolvidas para servirem a comunidade, aos outros desenvolvedores. Como exemplos citados por Pires (2013), o Facebook e o Twitter possuem APIs que possibilitam a interação com qualquer programador à sua base de dados, porém, de maneira limitada.

Além disso, as APIs podem ser ou não dependentes de uma linguagem de programação específica. Como um exemplo de não ser limitada somente a uma linguagem, temos a API do Google Maps, diferentes softwares desenvolvidos em diferentes linguagens conseguem usar a funcionalidade da dessa API.

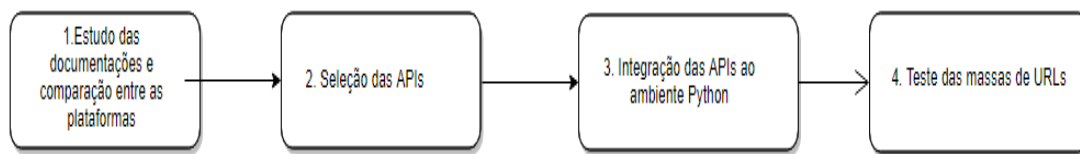
No contexto de dados abertos, as APIs são utilizadas para a disponibilização de dados acerca daquele portal, como por. exemplo, quantidade de datasets.

Algumas APIs, quando mais específicas, retornam informações dos próprios *datasets*, como por exemplo, data de criação, data da última modificação, nome, parâmetro de identificação, categorias.

3 MÉTODO

A metodologia do trabalho foi dividida em quatro partes. A Figura 1, mostrada logo abaixo, mostra resumidamente o que foi feito em cada uma delas e sua ordem cronológica até o fim do método desenvolvido:

Figura 1 – Metodologia implantada



Fonte: Elaborado pelo próprio autor (2019)

1. Na primeira etapa do projeto foi realizado um estudo da documentação das plataformas abertas disponíveis: foram avaliadas as documentações de oito plataformas e uma comparação foi realizada entre elas em termos de complexidade, quantidade de conteúdo, disponibilidade, documentação, APIs disponibilizadas resultando na Tabela 1, mostrada mais abaixo. As plataformas escolhidas foram:

- CKAN;
- ArcGis;
- Socrata;
- Junar;
- DKAN;
- OpenDataSoft;
- uData;
- Publish My Data

Tabela 1 – Comparação geral entre plataformas de dados abertos

	CKAN	ArcGis	Socrata	Junar	DKAN	ODS	PMMyData	uData
Categorização, apresentação visual	3	3	3	2	2	2	1	2
Análise de dados	2	3	3	2	3	2	1	2
<i>Open Source</i>	3	0	0	0	3	0	3	3
Disponibilização em versão gratuita	3	0	0	0	3	0	0	0
Documentação	3	1	3	3	3	3	1	1
Acesso aos dados via API	3	3	3	2	3	3	3	3
Variedade de APIs	3	2	2	2	2	3	2	1
Penetração de mercado	3	3	2	1	1	1	1	1

Fonte: Elaborado pelo próprio autor (2019)

Legenda: 3 – Muito bom/Existente. 2 – Bom. 1 – Razoável. 0 – Ruim/Inexistente.

Abaixo um detalhamento de cada um dos itens levados em consideração para o desenvolvimento da tabela comparativa.

- **Categorização e apresentação visual:** Nessa avaliação foram selecionados 10 portais para cada plataforma, com exceção das plataformas uData e Publish My Data por não terem sido identificados 10 portais governamentais que as utilizassem, e foram atribuídos valores de 0 a 3 conforme a disponibilização de termos pré-definidos para busca no portal, como por exemplo, agricultura, transporte, saúde, termos comuns e recorrentes em portais governamentais. Já o quesito de apresentação visual a avaliação teve o mesmo peso, de 0 a 3, avaliando a presença de elementos visuais que facilitassem a busca no portal, como por exemplo ícones identificando cada um dos grupos de *datasets*, como os já citados anteriormente, demonstrando características de acessibilidade ao público, algo muito importante. A tabela se encontra no Apêndice A e o cálculo para cada uma das plataformas foi um média dos valores atribuídos dividido pela sua quantidade, ou seja, para as plataformas com 10 portais o valor é a somatória das notas dividido por 20. Para arredondamento dos valores considerou-se a seguinte regra: notas finais com valor de 0 até 1,49 foram arredondadas para 1, de 1,5 até 2,49 foram arredondadas para 2 e notas de 2,5 até 3 foram arredondadas para 3.
- **Análise de dados:** Nesse quesito cada plataforma foi avaliada com relação à utilização, manipulação dos dados disponibilizados. As plataformas foram submetidas aos seguintes quesitos: gráficos e *dashboards* gerados, comparações de dados estatísticos.
- **Open Source:** Aqui foi feita uma avaliação direta e objetiva, se a plataforma disponibiliza seu código fonte de maneira aberta ou não.
- **Disponibilização em versão gratuita:** Aqui também uma avaliação direta e objetiva, caso a plataforma possua uma versão gratuita, seja a completa ou uma outra com limitações.
- **Documentação:** Avaliação feita em relação à volumetria de dados na documentação, à facilidade em encontrar a própria documentação, às vezes não tão exposta pelas plataformas, e as informações relacionadas às APIs.
- **Acesso aos dados via API:** Avaliação do processo de extração de dados via API.

- Variedade de APIs: Nesse quesito foi avaliada a quantidade de APIs por plataforma e a quantidade possível de informações diferentes a serem extraídas acerca de cada plataforma por suas APIs.
- Penetração de mercado: Avaliação feita com base nos resultados obtidos após o teste de identificação das plataformas, levou-se em conta a quantidade de portais governamentais que utilizam cada uma das plataformas em relação ao total de 1,379 URLs. De 0 a 10% do valor total foram avaliadas com valor 1, de 10,1 a 20% foram avaliadas com valor 2 e plataformas contendo de 20,1% até 100% foram avaliadas com valor 3.

2. Após o estudo da documentação das plataformas, começou-se a selecionar as APIs de cada uma delas, priorizando as que identificam unicamente o produto, isso porque em versões mais básicas dos algoritmos, diferentes assinaturas podem responder da mesma maneira, confundindo o desenvolvedor com relação a qual plataforma foi utilizada.

Terminado o estudo e leitura das documentações, as APIs foram selecionadas e algumas URLs submetidas a testes básicos na plataforma Postman para validá-las e certificar que estão promovendo um retorno correto e único. As APIs obtidas juntamente com seus respectivos exemplos de retorno, site da documentação e um exemplo para cada uma das plataformas foram especificados nos quadros a seguir:

CKAN	
DOCUMENTAÇÃO API	https://docs.ckan.org/en/2.8/api/
EXEMPLO DE PORTAL	Portal Brasileiro de dados abertos - http://dados.gov.br/
API IDENTIFICAÇÃO	GET /api/3/action/site_read
EXEMPLO RETORNO	<pre>{ "help": "http://dados.gov.br/api/3/action/help_show?name=site_read", "success": true, "result": true }</pre>
API COUNT DATASETS	GET /api/3/action/site_read
PALAVRA IDENTIFICAÇÃO COUNT DATASETS	COUNT
OBSERVAÇÕES	<p>A API utilizada para identificação das plataformas DKAN e CKAN é a mesma. Para diferenciá-las, posteriormente, foi utilizada a biblioteca Beautiful Soup, do Python, para ler o código –fonte HTML da página e procurar por pelo menos uma ocorrência da palavra ‘dkan’, e assim, concluindo a qual das duas plataformas o portal em questão faz uso.</p>

SOCRATA	
DOCUMENTAÇÃO API	https://dev.socrata.com/docs/endpoints.html
EXEMPLO DE PORTAL	Open Data Portal dataACT - https://www.data.act.gov.au
API IDENTIFICAÇÃO	GET /api/catalog/v1
EXEMPLO RETORNO	<pre>{ "results" : [{ "resource" : { "name" : "Lottery Cash 4 Life Winning Numbers: Beginning 2014", "id" : "kwxxv-fwze", "parent_fxf" : null, </pre>
API COUNT DATASETS	GET '/api/catalog/v1/domains?only=dataset&domains=' + domain + '&search_context=' + domain
PALAVRA IDENTIFICAÇÃO COUNT DATASETS	COUNT
OBSERVAÇÕES	

OPENDATASOFT	
DOCUMENTAÇÃO API	https://help.opendatasoft.com/apis/ods-search-v2/#search-api-v2
EXEMPLO DE PORTAL	Paris Data - https://opendata.paris.fr
API IDENTIFICAÇÃO	GET /api/v2
EXEMPLO RETORNO	<pre>{ "links": [{ "href": "https://opendata.paris.fr/api/v2", "rel": "self" }, </pre>
API COUNT DATASETS	GET /api/v2/catalog/datasets?rows=1
PALAVRA IDENTIFICAÇÃO COUNT DATASETS	TOTAL_COUNT
OBSERVAÇÕES	

ARCGIS	
DOCUMENTAÇÃO API	https://developers.ArcGis.com/documentation/core-concepts/rest-api/
EXEMPLO DE PORTAL	Portal da Prefeitura do Rio de Janeiro - https://www.data.rio/
API IDENTIFICAÇÃO	GET /api/v2
EXEMPLO RETORNO	<pre>{ "da-ta-sets": { "collection": "https://www.data.rio/api/v2/datasets?q={query} {&bbox,page[size,number],include,sort,filter,fields}", "object": "https://www.data.rio/api/v2/datasets/{:id}?{include}" }, }</pre>
API COUNT DATASETS	GET /data.json
PALAVRA IDENTIFICAÇÃO COUNT DATASETS	DATASET
OBSERVAÇÕES	

JUNAR	
DOCUMENTAÇÃO API	https://junar.github.io/docs/en/_sections/07-statistics.html#rest-endpoint
EXEMPLO DE PORTAL	Ministério de Economía y Finanzas del Perú - http://datosabiertos.mef.gob.pe
API IDENTIFICAÇÃO	GET /manageDeveloper/create
EXEMPLO RETORNO	<pre>{"pApiKey":"3f23dca5417790e25806a30d29ccaa27d1da4e81", "pPublicApiKey":"3d0154b610ab8e6df2aec884cf28e453acc299 91"}</pre>
API COUNT DATASETS	GET /api/v2/datasets/?auth_key=' + JUNAR_KEY + '&limit=1&format=json
PALAVRA IDENTIFICAÇÃO COUNT DATASETS	COUNT
OBSERVAÇÕES	<p>JUNAR_KEY: JUNAR_KEY é o valor da chave “pApiKey” gerado anteriormente na API de identificação da plataforma.</p> <p>Aqui há uma crítica específica com relação à plataformas Junar. Comparadas às outras, não disponibilizou, até onde o estudo foi feito, uma maneira de acessar as APIs apenas com a URL do portal como as outras, é preciso uma outra API disponibilizada para geração de <i>token</i>, identificado acima como “pApiKey” e adicioná-lo na URL para estar autenticado e autorizado a fazer a extração de dados.</p>

DKAN	
DOCUMENTAÇÃO API	https://docs.getdkan.com/en/latest/
EXEMPLO DE PORTAL	Portal de Dados Abertos da Califórnia - https://data.ca.gov/
API IDENTIFICAÇÃO	GET /api/3/action/site_read
EXEMPLO RETORNO	<pre>{ "help": "Return ``True``. :rtype: boolean", "success": true, "result": true }</pre>
API COUNT DATASETS	
PALAVRA IDENTIFICAÇÃO COUNT DATASETS	
OBSERVAÇÕES	<p>Para a plataforma DKAN, não foi encontrada uma API, até o momento do estudo, que disponibilizasse a quantidade de datasets em determinado portal. Mais uma vez a biblioteca Beautiful Soup foi utilizada. Durante os testes, percebeu-se que a quantidade de datasets disponibilizadas na página do site, em todas as urls DKAN testadas, ficavam armazenadas na mesma tag HTML, assim, no script foi adicionada um trecho de código para retornar o valor dessa tag como a quantidade de <i>datasets</i>.</p>

PUBLISH MY DATA	
DOCUMENTAÇÃO API	http://cofog01.data.scotland.gov.uk/documentation/api
EXEMPLO DE PORTAL	Portal Oficial de Estatísticas da Escócia - https://statistics.gov.scot/home
API IDENTIFICAÇÃO	GET /sparql.json?query=SELECT+%2A+WHERE+%7B%3Fs+%3Fp +%3Fo%7D+LIMIT+1
EXEMPLO RETORNO	{ "head": { "vars": ["s", "p", "o"]
API COUNT DATASETS	
PALAVRA IDENTIFICAÇÃO COUNT DATASETS	
OBSERVAÇÕES	Assim como o DKAN, também foi preciso utilizar o Beautiful Soup para extrair a quantidade de <i>datasets</i> de cada uma das urls.

UDATA	
DOCUMENTAÇÃO API	https://udata.readthedocs.io/en/stable/ https://dados.gov.pt/pt/apidoc/
EXEMPLO DE PORTAL	Portal de dados abertos da Administração pública – Portugal - https://dados.gov.pt/pt/
API IDENTIFICAÇÃO	GET /api/1/organizations/
EXEMPLO RETORNO	{ "data" : [{"acronym": "ama", "badges": [{"kind": "public-service"}, {"kind": "certified"}]},
API COUNT DATASETS	GET /api/1/datasets/?page_size=1
PALAVRA IDENTIFICAÇÃO COUNT DATASETS	TOTAL
OBSERVAÇÕES	

3. Nessa etapa as APIs foram integradas ao ambiente Python, utilizando a plataforma Jupyter Notebook, integração esta necessária para a execução dos algoritmos. Abaixo um exemplo do código para identificação das plataformas DKAN e CKAN e um exemplo do código para extração de *datasets* da plataforma DKAN. Em função da maioria das plataformas terem seus trechos de código muito parecidos, alterando apenas a API utilizada, somente essas duas plataformas tiveram seus códigos expostos.

DKAN/CKAN:

SCRIPT PARA IDENTIFICAÇÃO DAS PLATAFORMAS DKAN E CKAN

```
try:
    req = urllib.request.Request(
        url + dkan_ckan_sig,
        data=None,
        headers={
            'User-Agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_3)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/35.0.1916.47 Safari/537.36'
        }
    )
    resp = urllib.request.urlopen(req, timeout=20, context=ctx)
except: #raises HTTP, URL and refused connection errors
    dkan_error = True
    ckan_error = True
else:
    try:
        response_dict = json.loads(resp.read())
    except ValueError: #raises JSON format error of a non-expected result (API not
found)
        pass
    else:
        if 'help' in response_dict:
```

```

try:
    page = requests.get(domain)
    soup = BeautifulSoup(page.content, 'html.parser')
    the_word = 'dkan'
    words = soup.find(text=lambda text: text and the_word in text)
    catalog = 'CKAN'
    if words:
        catalog = 'DKAN'
except:
    dkan_error = True
    ckan_error = True

```

SCRIPT PARA EXTRAÇÃO DA QUANTIDADE DE DATASETS

```

if row_url['PRODUCT'] == 'DKAN':
    try:
        url = row_url['URL'] + '/search'

        req = urllib.request.Request(
            url,
            data=None,
            headers={
                'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/71.0.3578.98 Safari/537.36'
            }
        )
        resp = urllib.request.urlopen(req, timeout=60, context=ctx)

    except Exception as e: #raises HTTP, URL and refused connection errors
        print(str(row_url['ID']) + ' ' + url + ' CONNECTION ISSUE\n')
        file_errors.write(str(datetime.datetime.now()) + '|' + str(row_url['DEPTH']) + '|' +

```

```

str(row_url['ID']) + '|' + str(row_url['URL']) + '|' + str(row_url['DOMAIN']) + '|DKAN|' +
str(e) + '\n')

    else:
        #parse JSON
        try:
            page = requests.get(url)
            soup = BeautifulSoup(page.content, 'html.parser')
            resultado = soup.find("div", class_="view-header")

        except (ValueError, TypeError, KeyError) as e:
            print(str(row_url['ID']) + ' ' + url + ' ' + 'PRODUCT:ERROR/UNKNOWN\n')
            file_errors.write(str(datetime.datetime.now()) + '|' + str(row_url['DEPTH']) + '|'
+ str(row_url['ID']) + '|' + str(row_url['URL']) + '|' + str(row_url['DOMAIN']) + '|DKAN|'
+ str(e) + '\n')

        else:
            print(str(row_url['ID']) + ' ' + url + ' ' + (str(resultado.text)) + '\n')
            file_results.write(str(datetime.datetime.now()) + '|' + str(row_url['DEPTH']) + '|'
+ str(row_url['ID']) + '|' + str(row_url['URL']) + '|' + str(row_url['DOMAIN']) + '|DKAN|'
+ str(resultado.text) + '|' + domain_country + '|' + ip_country + '|' + root_domain + '\n')

```

4. Como última etapa do projeto testou-se as massas de URLs a fim de identificar a assinatura do portal utilizado, caso tenha utilizado alguma, e promover a extração da quantidade de *datasets* de cada um deles. As URLs foram obtidas a partir do trabalho de Santos (2018) em que possíveis URLs de portais de dados abertos eram catalogadas a partir da base de dados *Common Crawl*, um projeto responsável por indexar dados de toda a internet.

Os testes foram realizados em uma máquina virtual Windows, o script Python foi executado pelo terminal e os resultados eram escritos em arquivos separados, tanto as execuções de erro, como as URLs que não pertenciam a nenhuma das plataformas, como as que foram identificadas.

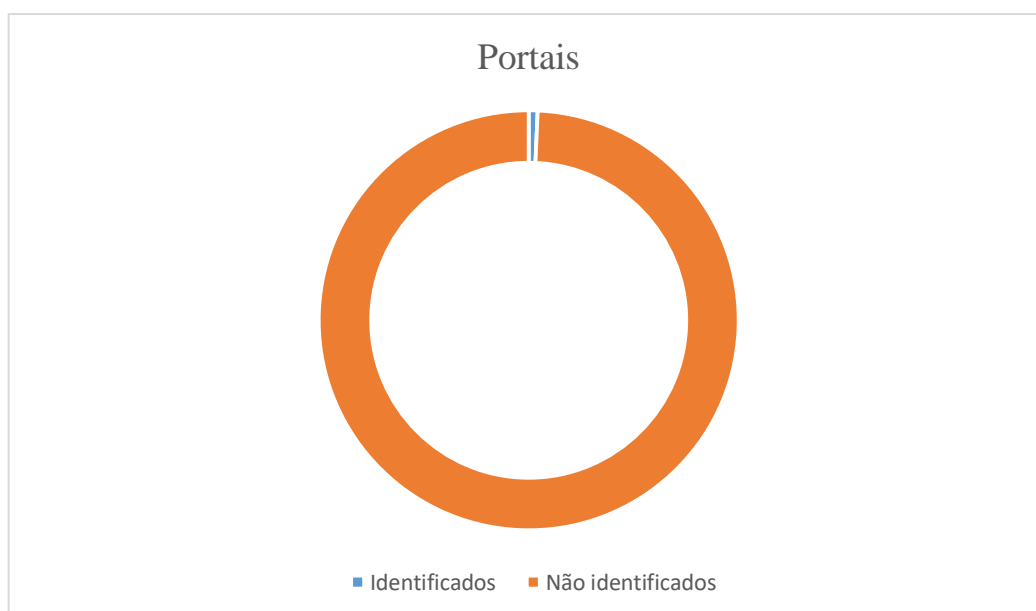
Todos os testes e seus dados obtidos estão descritos na sessão Resultados.

4 RESULTADOS

1º Teste

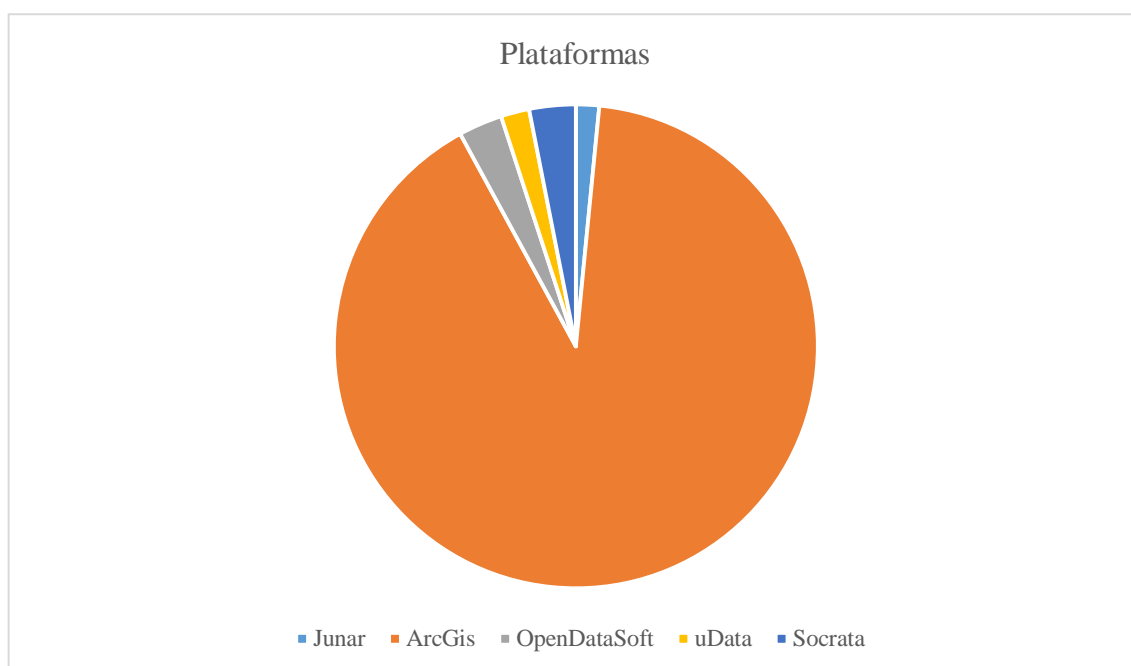
Nesse primeiro teste foram disponibilizadas 85.032 URLs para o script de verificação e como resultado foram identificados 580 portais. A figura 2 mostra a quantidade de portais que foram identificados como sendo de uma das plataformas escolhidas no trabalho. Já a Figura 3 evidencia quais plataformas foram identificadas e a quantidade de cada uma delas.

Figura 2 – Quantidade de portais identificados



Fonte: Elaborado pelo próprio autor (2019)

IDENTIFICADOS:580, NÃO IDENTIFICADOS: 84452

Figura 3 – Quantidade de portais identificados por plataforma

Fonte: Elaborado pelo próprio autor (2019)

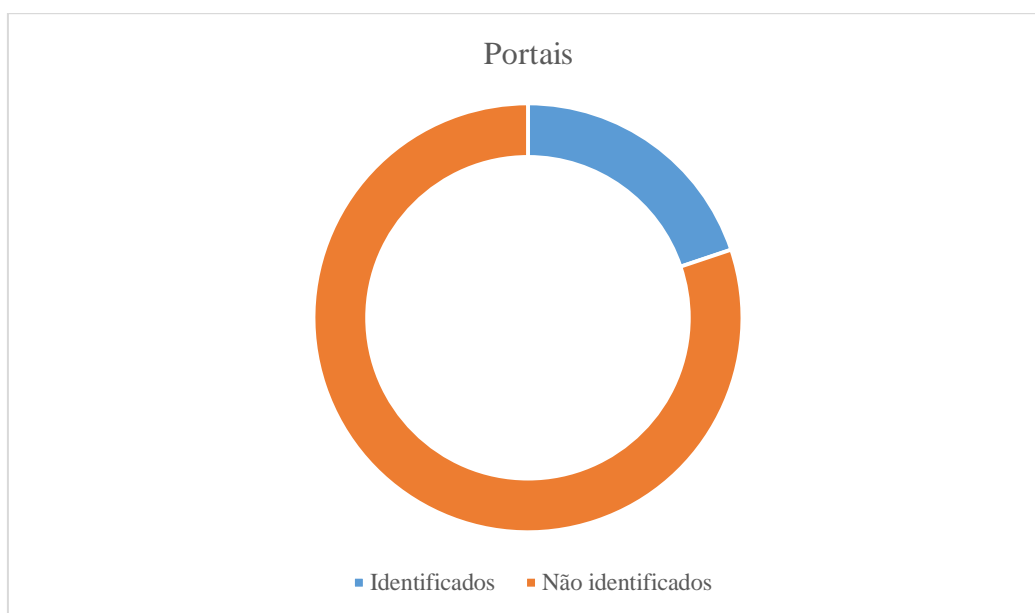
JUNAR:9, ARCGIS: 525, OPENDATASOFT:17, UDATA:11, SOCRATA:18

Nesse primeiro teste há uma discrepância muito grande entre a quantidade de plataformas. Após perceber os resultados, uma busca pela causa raiz dessa diferença foi iniciada. Depois da análise, identificou-se que a massa utilizada para teste era em sua maioria composta por URLs contendo em seu próprio domínio a palavra “arcgis”, e assim, pertencendo a uma única plataforma.

2º Teste

Em função do pouco número de plataformas identificadas no teste anterior, realizou-se um novo teste com uma nova carga de URLs e, apesar de um número significativamente menor, foi possível encontrar registros de todas as plataformas escolhidas para o trabalho. Foram submetidas ao script de identificação um total de 4.025 URLs, sendo identificados 799 portais. A figura 4 mostra a quantidade de portais que foram identificados como sendo de uma das plataformas escolhidas no trabalho. Já a Figura 5 apresenta a distribuição desses portais nas plataformas.

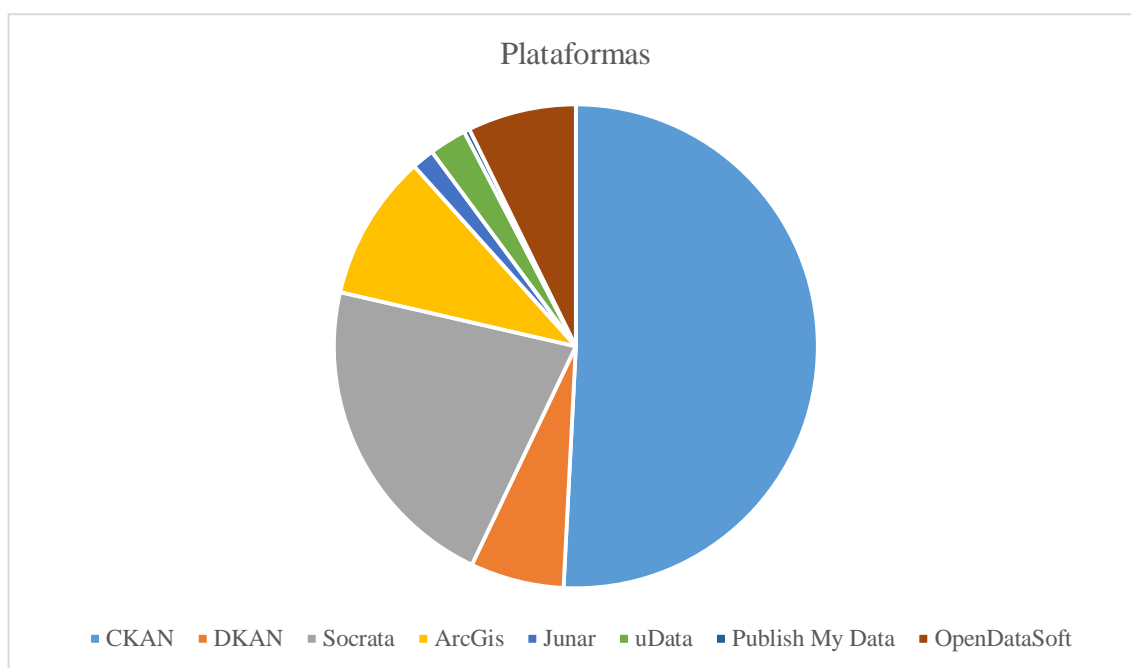
Figura 4 – Quantidade de portais identificados



Fonte: Elaborado pelo próprio autor (2019)

IDENTIFICADOS:799, NÃO IDENTIFICADOS: 3226

Figura 5 – Quantidade de portais identificados por plataforma



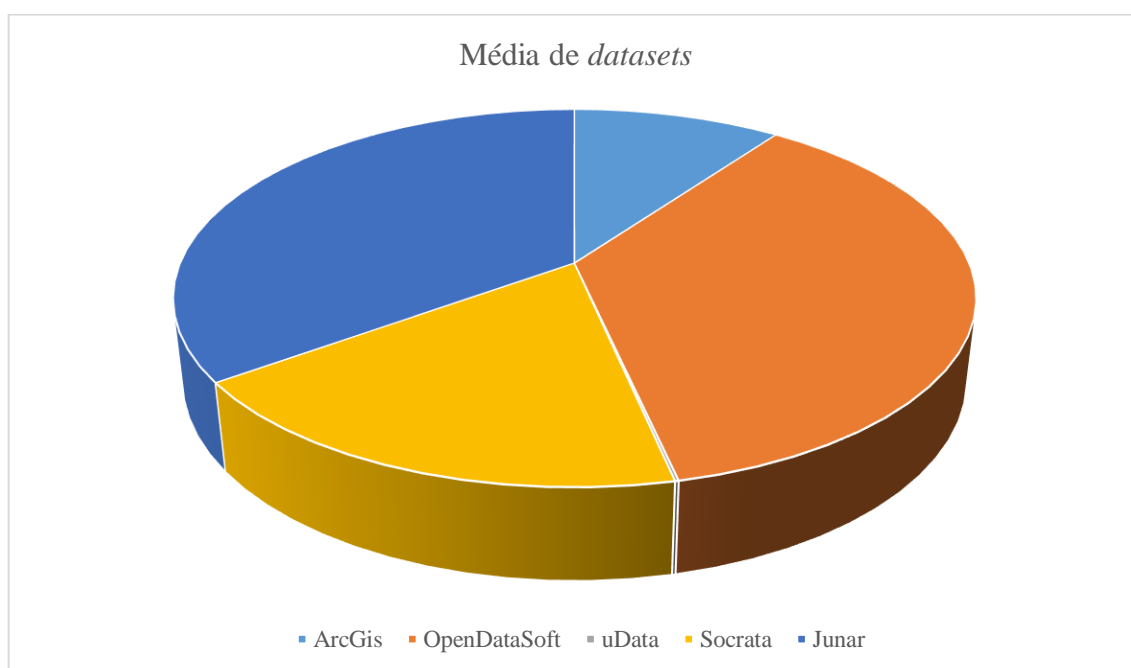
Fonte: Elaborado pelo próprio autor (2019)

CKAN: 406, DKAN:50, SOCRATA:172, ARCGIS:78, JUNAR: 12, UDATA:20, PUBLISH MY
DATA: 3, OPENDATASOFT:58

3º Teste - Extração de *datasets*

Nesta etapa de testes, os portais identificados pelo 1º teste foram submetidos ao script de extração da quantidade de *datasets* e na Figura 6 foi exposto para cada uma das plataformas encontradas a média de *datasets* por portal, ou seja, dividiu-se a quantidade total de *datasets* retornados pelo número de portais identificados

Figura 6 – Média de *datasets* por plataforma identificada

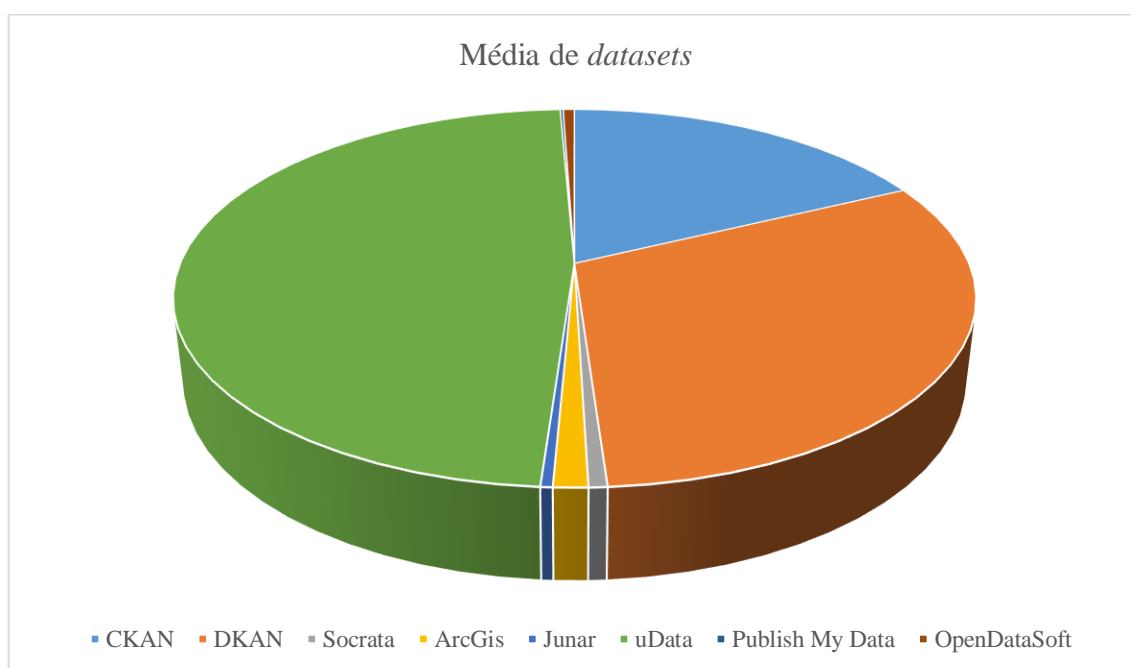


Fonte: Elaborado pelo próprio autor (2019)

ARCGIS: 82 OPENDATASOFT: 300 UDATA: 1 SOCRATA:151 JUNAR:287

Já no teste exposto na Figura 7, os valores também são correspondentes às médias de *datasets* retornados por cada uma das plataformas após o script de extração de quantidade de *datasets*, porém, agora, para as URLs retornadas no 2º teste.

Figura 7 – Média de *datasets* por plataforma identificada



Fonte: Elaborado pelo próprio autor (2019)

CKAN:14340 DKAN: 24870 SOCRATA: 507 ARCGIS: 943 JUNAR: 331 UDATA: 38643

PUBLISH MY DATA: 113 OPENDATASOFT: 432

5 CONSIDERAÇÕES FINAIS

5.1 Conclusão

Oito plataformas foram selecionadas como as principais para a realização do trabalho. A quantidade de artigos e trabalhos foi fundamental para a escolha, visto que estas foram as mais citadas ao longo da busca pela internet: CKAN, Socrata, ArcGis, OpenDataSoft, Publish My Data, DKAN, Junar e uData.

O estudo da documentação das plataformas foi realizado e cada semelhança entre as plataformas era documentada e serviu como base para a criação da tabela comparativa entre elas. Dados sobre a documentação, sobre a usabilidade dos portais, geração e conteúdo e, posteriormente, os resultados dos scripts foram todos considerados. Paralelamente, na leitura das documentações, as APIs disponibilizadas por cada plataforma eram analisadas e ao final, foram escolhidas as que identificariam unicamente o produto, com exceção da dupla CKAN/DKAN que necessitou das funcionalidades da biblioteca Beautiful Soup para diferenciar uma da outra.

Com todas as APIs escolhidas, o desenvolvimento dos algoritmos responsáveis pela identificação das plataformas foi desenvolvido, em Python, e os testes começaram. Foram realizados testes com duas massas de URLs em função de uma delas não conter todas as plataformas, concentrando a maioria dos portais em uma plataforma só (ArcGis).

Ao término do processamento das 90.000 URLs, número escolhido para testes, aquelas que foram identificadas com algum dos portais escolhidos no trabalho (cerca de 1.400 URLs no total, incluindo os dois testes) tiveram seus dados armazenados em um arquivo formato .csv e foram submetidas ao segundo script para verificação da quantidade de *datasets*.

5.2 Trabalhos futuros

Até o momento presente, o trabalho conta com oito plataformas identificadas e com seus dados de *datasets* extraídos pelos scripts. Tem-se como trabalho futuro a identificação de mais plataformas e testes com massas de testes maiores, com cerca de 1.000.000 de URLs. Além disso, pretende-se promover uma comparação mais ampla entre elas, com mais itens a serem analisados.

6 REFERÊNCIAS

APPLICATION programming interface. 2015. Disponível em: <https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/WIKIPEDI/W120623A.pdf>. Acesso em: 14 maio 2016.

PORTAL BRASILEIRO DE DADOS ABERTOS (Brasil.). Aplicativos e serviços que utilizam dados abertos. [2015]. Disponível em: <http://dados.gov.br/pagina/aplicativos>. Acesso em: 15 nov 2018.

CONEGLIAN, C. S; SEGUNDO, J. E. S; Profissional da Informação no contexto dos dados abertos: uso do CKAN para a disponibilização e a organização de dados. V. 5, n. 2. 2016.

CORRÊA, A. S; ZANDER, Pär-Ola; SILVA, F. S. C. da. Investigating open data portals automatically: a methodology and some illustrations. 2018. Disponível em: http://delivery.acm.org/10.1145/3210000/3209292/a82-correa.pdf?ip=179.159.215.23&id=3209292&acc=AUTHOR%2DIZED&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E08F9E9F457215076&__acm__=1572392159_1a2b669185ad68148e2f61a5c3b122cb

CUNHA, M. A. V. C. da; ROSINA, M. S. G; TEIXEIRA, M. A. C; SILVA, A. P. da; LAZZARI, E. A; SILVA, M. C. F. da; KAROLCZAK, R. M; CANTONI, S. L; COELHO, T. R; SILVA, T. A. B; SPINOLA, Larissa; MARINHO, Lucas; SCHELIGA, N. R. Dados abertos nos municípios, estados e governo federal brasileiros. Technical Report. Democracia Digital. 2015

OPEN Data Barometer. Global Report. Fourth Edition. Technical Report. The World Wide Web Foundation. Disponível em: <http://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf>). Acesso em: 18 maio 2018.

BATISTA, A. H. Portal Brasileiro de Dados Abertos. 2015. Disponível em: <http://www.cgu.gov.br/sobre/institucional/concursos/concurso-de-boas-praticas/arquivos/3o-concurso-de-boas-praticas/og-05-mpog.pdf>). Acesso em: 14 junho 2018.

ESRI. An Open Data Approach That Works. 2015. Disponível em: <http://www.esri.com/library/brochures/pdfs/open-data-government.pdf>. Acesso em 02 out 2018.

HERRMANN, Augusto. Minicurso de CKAN. LOD Brasil Linked Open Data. Disponível em: <http://www.inf.ufsc.br/~jose.todesco/LODBrasil/Minicurso/CKAN.pdf>. Acesso em 02 out 2018.

ISOTANI, Seiji; BITTENCOURT, I. I. Dados Abertos Conectados. 2018. Disponível em: <http://www.pgcl.uenf.br/2018/download/DadosAbertosConectados.pdf>. Acesso em 01 out 2018.

JUNAR. Junar Open Data Platform. 2014. Disponível em: <http://junar-cdnbrandings.s3.amazonaws.com/reference-material/Product-sheet-Open-Data-Portal-N0515.pdf>. Acesso em 02 out 2018.

OPEN Data Handbook. Documentation: Release 1.0.0. 2012. Open Knowledge Foundation. Disponível em: http://www.edinburgh.gov.uk/download/downloads/id/3392/open_data_handbook.pdf. Acesso em: 14 nov 2018.

SANTOS, Alisson Lúcio dos. Método de catalogação de portais de dados abertos a partir de URLs indexadas no Common Crawl. 2018. Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Campinas, 2018.

UBALDI, Barbara. 2013. Open Government Data: towards empirical analysis of open Government Data Initiatives. OECD Working Papers on Public Governance. Organization for Economic Co-operation and Development, Paris. Retrieved 2014-11-18 from <http://dx.doi.org/10.1787/5k46bj4f03s7-en>.

UNITED STATES OF AMERICA. DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS. (Org.). United Nations: E-Government Survey 2016. 2016. Disponível em: <http://workspace.unpan.org/sites/Internet/Documents/UNPAN97453.pdf>. Acesso em: 14 maio 2018.

APÊNDICE A – Tabela comparativa entre as plataformas acerca de categorização e apresentação visual

	Plataforma	Categorização	Apresentação visual
dados.recife.pe.gov.br	CKAN	3	3
dados.tce.rs.gov.br	CKAN	3	3
data.qld.gov.au	CKAN	3	3
dati.toscana.it	CKAN	3	0
bermuda.io	CKAN	3	0
data.gov.ro	CKAN	3	3
dados.rs.gov.br	CKAN	3	3
data.salzburgerland.com	CKAN	3	0
data.gov.bf	CKAN	3	3
data.gov.gr	CKAN	3	3
www.ctdatacollaborative.org	DKAN	3	2
schoolgrades.georgia.gov	DKAN	2	0
docs.digital.mass.gov	DKAN	2	0
data.adb.org	DKAN	3	0
gishubdata.nd.gov	DKAN	3	3
healthdata.gov	DKAN	3	2
data.louisvilleky.gov	DKAN	3	3
research.mdc.mo.gov	DKAN	2	0

dati.gov.it	DKAN	3	3
data.gov.ph	DKAN	3	3
www.data.rio	ARCGIS	3	3
opendata.minneapolismn.gov	ARCGIS	3	3
data.cityofpasadena.net	ARCGIS	3	3
data.cityofsacramento.org	ARCGIS	3	3
opendata.dc.gov	ARCGIS	3	3
datosabiertos.esri.co	ARCGIS	3	3
data.cityofsydney.nsw.gov.au	ARCGIS	3	3
opendata.atlantaregional.com	ARCGIS	3	3
gis- michigan.opendata.arcgis.com	ARCGIS	3	3
datos- geored.opendata.arcgis.com	ARCGIS	3	3
opendata.terrassa.cat	SOCRATA	3	3
data.bathhacked.org	SOCRATA	3	3
data.chattlibrary.org	SOCRATA	3	2
data.cityofchicago.org	SOCRATA	3	2
data.honolulu.gov	SOCRATA	3	3
data.austintexas.gov	SOCRATA	3	3
data.act.gov.au	SOCRATA	3	3
data.providenceri.gov	SOCRATA	3	3
data.weho.org	SOCRATA	3	3

data.colorado.gov	SOCRATA	3	3
ressources.data.sncf.com	ODSOFT	3	0
data.angers.fr	ODSOFT	3	3
observa.gijon.es	ODSOFT	3	3
opendata.bristol.gov.uk	ODSOFT	3	0
data.loire-atlantique.fr	ODSOFT	3	3
opendata.brussel.be	ODSOFT	3	0
opendata.paris.fr	ODSOFT	3	0
data.iledefrance.fr	ODSOFT	3	2
parisdata.opendatasoft.com	ODSOFT	3	2
data.eindhoven.nl	ODSOFT	3	3
datos.arsat.com.ar	JUNAR	3	3
datos.sanlorenzo.com.ar	JUNAR	3	3
datosabiertos.enacom.gob.ar	JUNAR	3	3
datos.villamaria.gob.ar	JUNAR	3	0
escazu.opendata.junar.com	JUNAR	3	3
vicentelopez.opendata.junar.com	JUNAR	3	3
defensoriacr.opendata.junar.com	JUNAR	3	3
datosabiertos.olavarria.gov.ar	JUNAR	3	0

datos.mpuentealto.cl	JUNAR	3	0
cityofpaloalto.org	JUNAR	3	0
forms.datalynk.ca	UDATA	-	-
data.gouv.fr	UDATA	3	0
datingchile.cl	UDATA	-	-
intheclosetdating.com	UDATA	-	-
jewishdatingsingles.com	UDATA	-	-
freedatingforyou.com	UDATA	-	-
heavymetaldating.com	UDATA	-	-
cyprusdating4u.com	UDATA	-	-
dataclustersystem.com	UDATA	-	-
emailer.datafeedfile.com	UDATA	-	-
statistics.data.gov.uk	PMYDATA	2	0
opendatacommunities.org	PMYDAT	2	0