

Prémio Arquivo.pt

Descrição Sumária do Trabalho

Identificação

- Título: Pegada Lusa
- Área temática: Sustentabilidade, Tecnologia, Ciência de Dados, Ciências Sociais
- Candidatos: Diana Costa & Sérgio Teixeira
- Email:
 - dianacosta87@gmail.com
 - sergioramosteixeira@gmail.com
- Website: pegadalusa.pt

Descrição do Trabalho

O projeto "Pegada Lusa" surge como uma iniciativa tecnológica de análise de dados para mapeamento da sustentabilidade em Portugal. Com o suporte do Arquivo.pt, a "Pegada Lusa" analisa milhares de informações nacionais e locais relatar a evolução das políticas e iniciativas sustentáveis por região. Para essa análise são considerados os projetos e boas práticas englobados nos Objetivos de Desenvolvimento Sustentável (ODS), definidos pelas Nações Unidas para a Agenda 2030, e notícias relacionadas com sustentabilidade nos jornais locais de cada região.

Além da componente jornalística e dos ODS, a "Pegada Lusa" oferece uma ferramenta de análise da evolução de preços entre meios de transporte, comparando-os com as emissões produzidas.

A elaboração deste projeto foi compreendida em várias fases:

1. Pesquisa dos dados necessários e fontes de informação elegíveis pelo Arquivo.pt para a obtenção do histórico desses mesmos dados.

Alguns dos dados que eram pretendidos para o desenvolvimento deste projeto encontravam-se em fontes de informação que não estavam disponibilizadas no Arquivo.pt, ou que o seu conteúdo se encontrava em branco, por estas fontes terem mecanismos contra *robots*. Assim sendo, esta fase acabou por ter um peso mais elevado do que o imaginado, para que fosse possível encontrar fontes de informação completamente elegíveis pelo Arquivo.pt.

2. Extração dos diversos dados das próprias páginas web para dados atuais e do Arquivo.pt para os dados históricos.

Para a extração dos dados foram criados e usados algoritmos em *Python* para o *web scrapping*. Teve de ser criado um algoritmo para cada fonte de dados, para se obter a informação pretendida, uma vez que eram necessários diversos tipos de dados de diversas fontes. Na tabela abaixo, pode-se observar as fontes usadas para cada tipo de informação.

Tipo de Dados	Dados Atuais	Dados Históricos
ODS - Projetos	odslocal.pt	odslocal.pt via arquivo.pt
ODS – Boas Práticas	odslocal.pt	odslocal.pt via arquivo.pt
Transportes – Combustíveis	maisgasolina.com	maisgasolina.com
Transportes – Portagens	brisaconcessao.pt	brisa.pt e brisaconcecao.pt via arquivo.pt
Transportes – Comboios	cp.pt	cp.pt via arquivo.pt
Transportes – Bus	viarail.pt	viarail.pt e checkmybus.pt via arquivo.pt
Transportes – Aviões	kiwi.com e omio.com	edreams.pt, rumbo.pt e kiwi.com via arquivo.pt

Para a extração de notícias, todos os dados foram recolhidos do Arquivo.pt, através da sua API, onde eram efetuadas pesquisas através das palavras-chave ligadas à sustentabilidade e uma lista de fontes, com todos os jornais e revistas locais, que foi obtida do Portal de dados abertos da Administração Pública (dados.gov.pt). Aqui, foi usada a biblioteca *Newspaper3k* que auxilia ao *web scrapping* de notícias, identificando as diversas componentes da notícia, sem a necessidade de um algoritmo individual para cada fonte de dados.

3. Processamento de informação para obtenção dos dados necessários e inserção na base de dados.

Por vezes, as notícias podem ter o termo pesquisado no texto sem que estas se encontrassem diretamente associadas à sustentabilidade, para considerar apenas as notícias relevantes, foi atribuído um peso a cada um dos termos, e apenas eram consideradas as notícias cuja soma de termos multiplicados pelo seu peso atingissem determinado valor.

Além disso, foi efetuada uma análise de sentimento através de PLN (Processamento de Linguagem Natural), para interpretar se a notícia é positiva ou negativa de acordo com a sua informação, ou seja, se uma notícia indica que a poluição aumentou, o sentimento será negativo, enquanto uma notícia que indique que foram efetuadas medidas para combater a poluição, terá um sentimento positivo. Uma vez que esta análise é efetuada através de um processamento automático, não está livre de erros – através de uma análise a um universo de 50 notícias, foi verificada uma taxa de erro inferior a 10%.

Após a extração e seu processamento, todos os dados necessários foram introduzidos na base de dados na nuvem *Cloud Firestore* do *Google Firebase*.

4. Desenvolvimento do website para disponibilização da informação.

Esta é a parte visível para o utilizador, onde através de um *design* intuitivo e responsivo, o utilizador pode navegar entre as diversas informações disponibilizadas e verificar factos atuais e históricos sobre sustentabilidade em Portugal e nos seus distritos.

O website disponibiliza na primeira página um mapa de Portugal interativo, onde o utilizador pode selecionar o distrito a visualizar, como mostra a imagem abaixo.

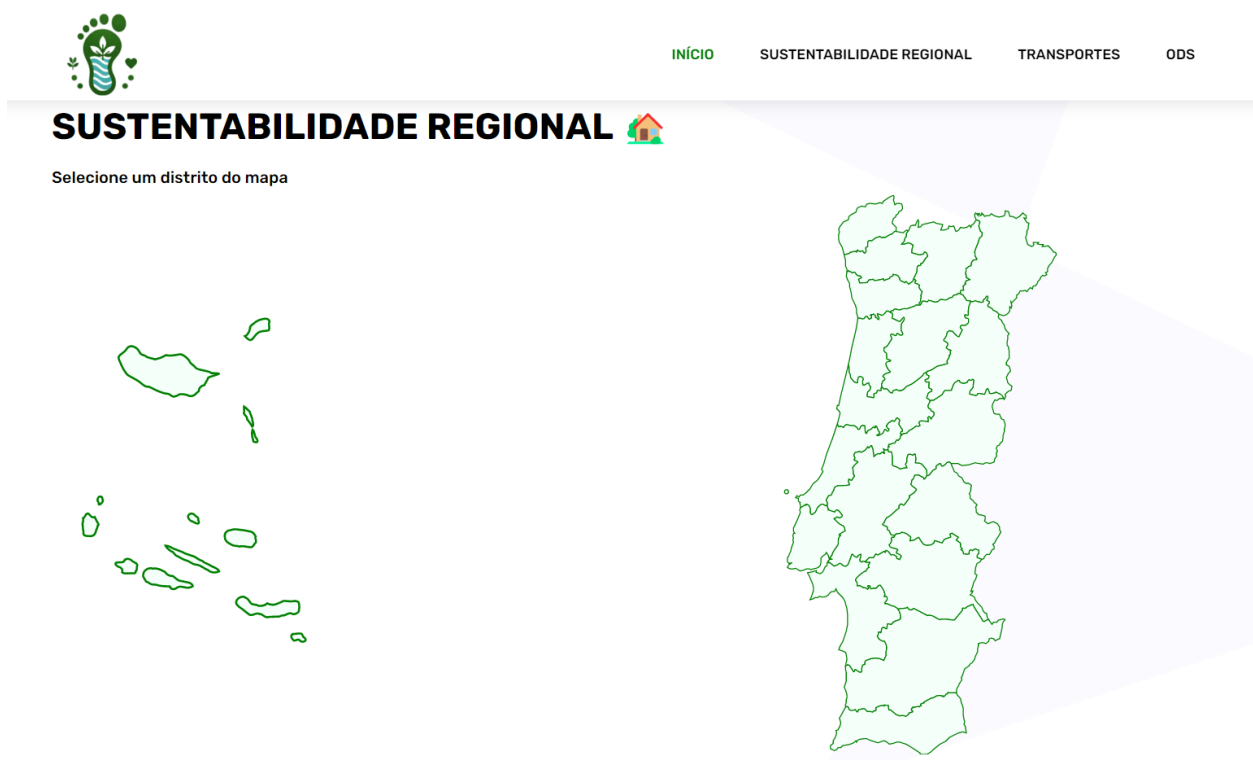


Figura 1 - Mapa interativo de escolha do distrito

Dentro da página do distrito é possível observar no topo um resumo dos dados analisados, incluindo o *GreenScore* – nome criado para o índice qualificativo de sustentabilidade baseado na análise sentimental das notícias, o número de notícias analisadas no Arquivo.pt, ou seja, que foram encontradas na pesquisa de termos, e o número de notícias relevantes, cujo algoritmo identificou que realmente estavam relacionadas com sustentabilidade. É ainda apresentado o número de projetos e boas práticas ODS em atividade.

De seguida, são mostradas as notícias relevantes, assim como a lista de fontes de informação usadas na pesquisa, e um *WordCloud* com os termos mais usados nas notícias. É possível clicar nos termos e nas fontes de informação e filtrar as notícias por essas entidades.



Figura 2 - Notícias locais, termos mais usados e fontes de informação usadas.

Ainda na página do distrito, é possível observar a evolução ao longo dos anos do número de iniciativas para cada objetivo dos ODS, assim como verificar a lista de boas práticas implementadas desde o início dos ODS em 2020, até ao dia de hoje, sendo possível filtrar as iniciativas que já foram descontinuadas e as que continuam ativas, os objetivos associados à iniciativa e o ano.



FILTROS:

Todos

No Ativo

Descontinuados

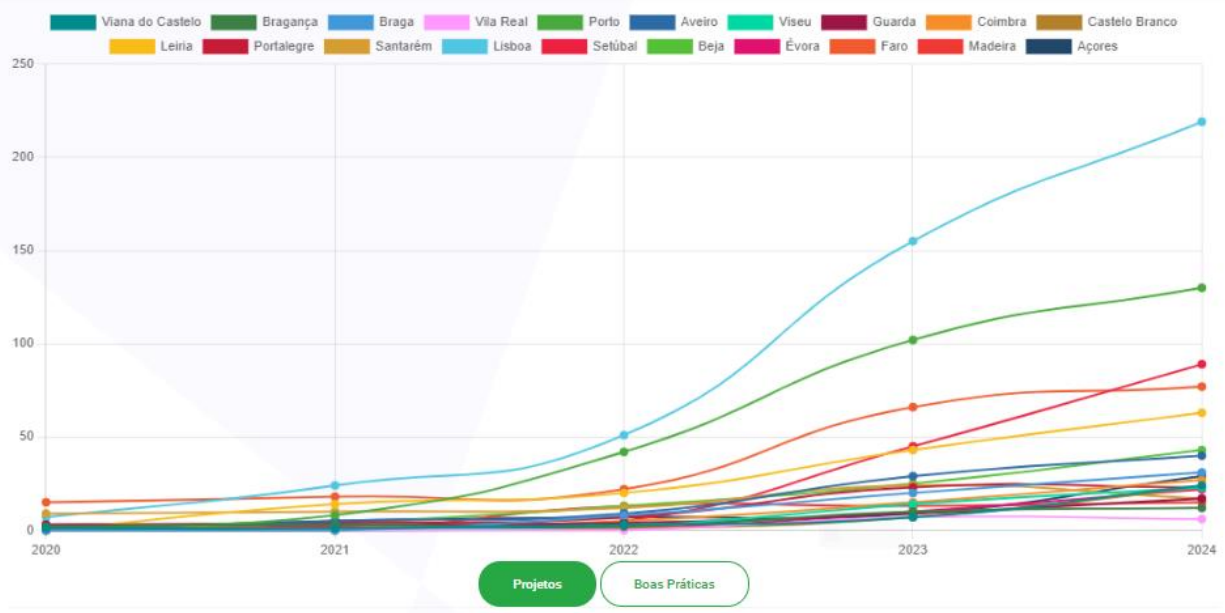
Objetivos

Anos

INICIATIVA	OBJETIVOS	2020	2021	2022	2023	2024
 Eu compro local Funchal	 	×	×	✓	×	×
Estratégia municipal para a Educação Ambiental e Desenvolvimento Sustentável Funchal	 	×	×	✓	×	×
 PEAS - Programa de Educação Ambiental para a Sustentabilidade Porto Santo	 	×	×	✓	✓	✓
 Sustentabilidade nos Mercados Municipais Funchal	 	×	×	✓	×	✓
 Balcão do Investidor Funchal	 	×	×	×	×	✓
Limpeza das linhas de água e de sarjetas Funchal	 	×	×	✓	✓	✓
 Funchal Alerta	 					

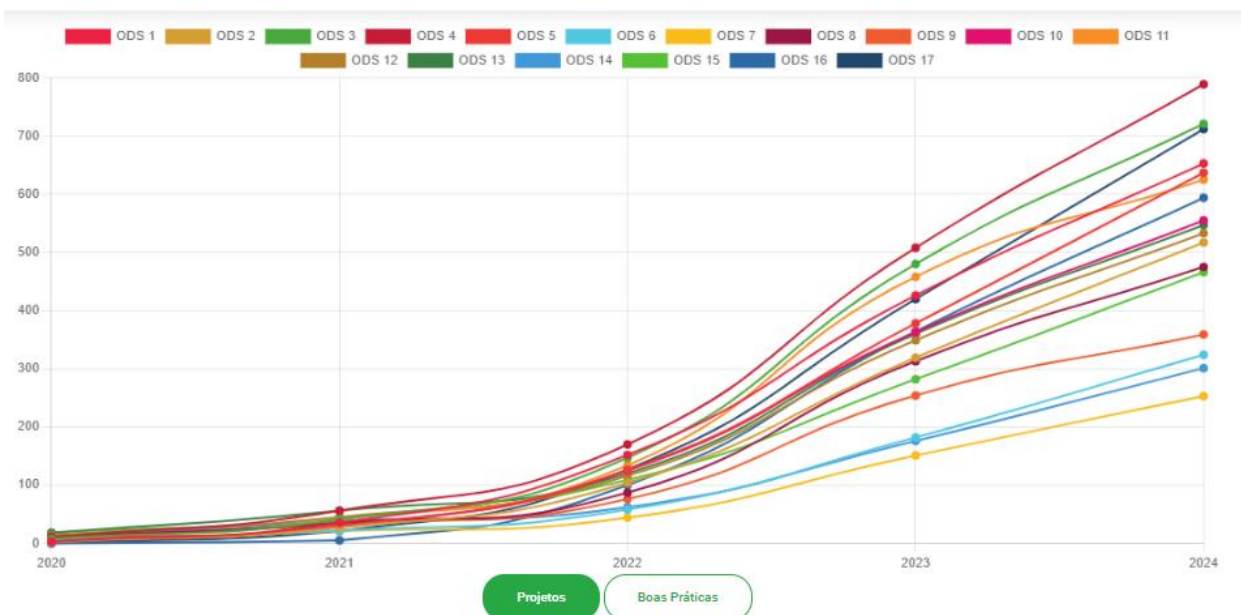
Figura 3 - Boas Práticas no âmbito dos ODS

Foi criada também uma outra página dedicada aos Objetivos de Desenvolvimento Sustentável, mas a nível nacional, no qual é possível observar a evolução do número de projetos e boas práticas por cada um dos dezassete objetivos ao longo dos anos, assim como, a evolução do total dessas iniciativas implementadas por distrito. Ambas as informações são também representadas graficamente para uma melhor interpretação e comparação dos dados.



DISTRITO		2020	2021	2022	2023	2024
 VIANA DO CASTELO	Projetos	1	1	3	7	24
	Boas Práticas	2	3	8	16	167
 BRAGANÇA	Projetos	2	3	2	10	12
	Boas Práticas	2	2	3	9	10

Figura 4 - Evolução dos ODS por distrito



ODS		2020	2021	2022	2023	2024
	Projetos	45	158	262	721	1041
	Boas Práticas	42	106	250	774	2012
	Projetos	3	35	152	426	653
	Boas Práticas	3	12	33	125	319
	Projetos	2	27	104	319	517

Figura 5 - Evolução dos ODS por objetivo, a nível nacional

O projeto oferece ainda uma ferramenta de análise da evolução dos preços dos transportes, comparando com as emissões da viagem. Para esta análise, é tido em conta o itinerário Porto – Lisboa, uma vez que este trajeto liga as duas maiores cidades portuguesas, e ambas possuem aeroporto, central de camionagem e estação ferroviária, possibilitando a comparação de vários meios de transporte.

Nesta ferramenta, o utilizador pode escolher de 1 a 5 passageiros e comparar o preço total dessa viagem ao longo dos anos nos diversos meios de transporte (carro, autocarro, comboio e avião), e qual a pegada resultante das emissões de carbono. Além disso, o utilizador pode escolher o tipo de combustível, a categoria de portagem, a classe (Turística e Executiva) e o comboio (InterCidades e AlfaPendular), e para os autocarros e aviões pode escolher entre viagem planeada ou de última hora – uma vez que estes preços são voláteis.

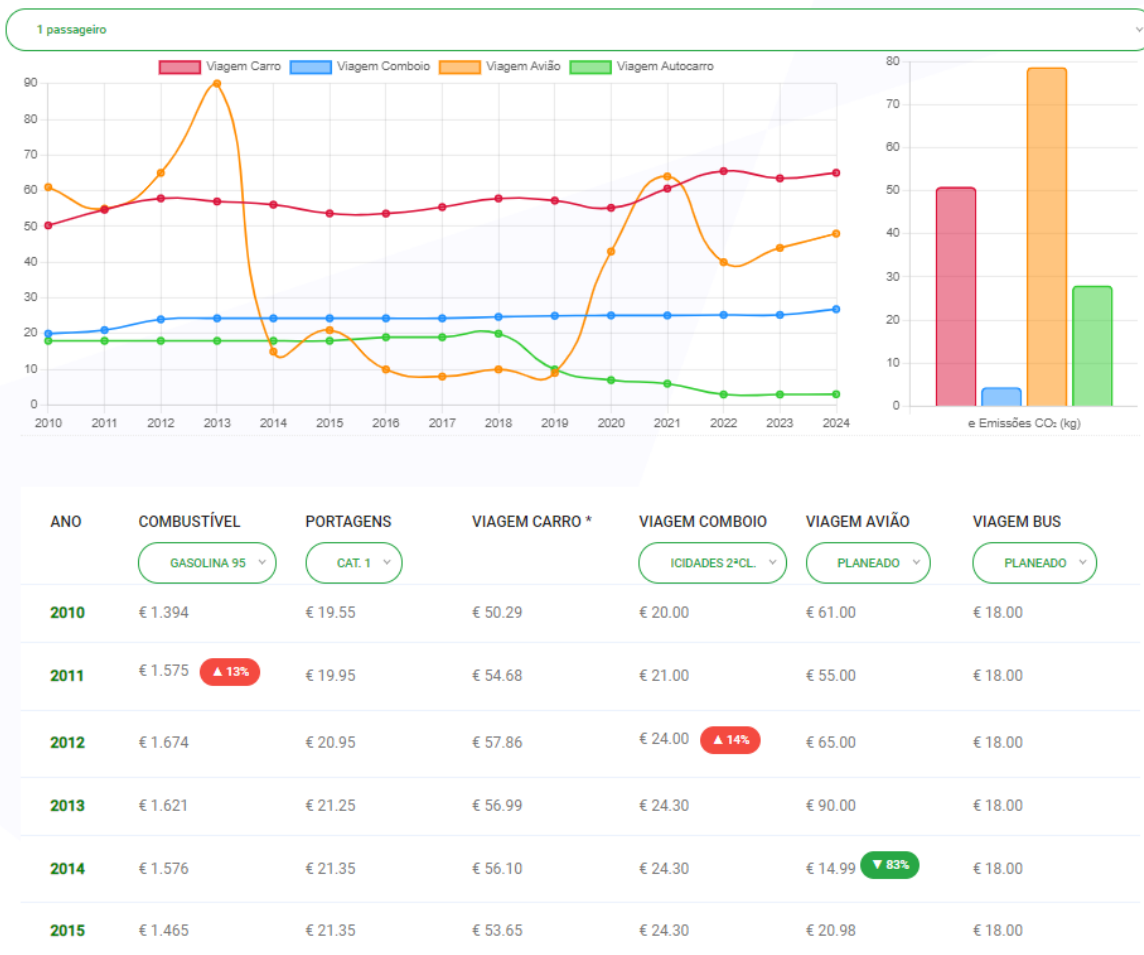


Figura 6 - Ferramenta de comparação de preços e emissões dos transportes.

O valor usado na ferramenta é o mais baixo encontrado no Arquivo.pt, porém, pode ter existido preços mais baixos em dias que a fonte não tenha sido arquivada, uma vez que os preços dos combustíveis, do avião e autocarro (desde 2019) são voláteis.

O objetivo desta ferramenta é analisar se estão a ser tomadas medidas para incentivo do uso de transportes menos poluentes, o que não se verifica, uma vez que o comboio (menos poluente) tem tido um aumento constante do preço, ao contrário do autocarro que baixou bastante os preços com o aumento da concorrência. Comparando com o carro, a viagem de comboio só compensa com três ou mais passageiros, tal e qual como em 2010, não sendo registada uma evolução ecológica neste aspeto.

Objetivos

Numa altura em que todos os dias se fala na crise climática, procuramos criar uma plataforma que pudesse informar sobre as práticas de sustentabilidade em Portugal, por região, e como

têm contribuído positiva ou negativamente para a proteção ambiental e promoção da sustentabilidade. O grande objetivo deste projeto não é ser apenas uma ferramenta de análise e mapeamento, mas também uma iniciativa estratégica para impulsionar a consciência ambiental e catalisar mudanças significativas.

Resultados Atingidos

Para a realização deste trabalho foi efetuado o *web scraping* a 451732 artigos de jornais locais, através do Arquivo.pt, considerando 15916 desses artigos como relevantes e, por isso, armazenados na base de dados da Pegada Lusa. Foram recolhidos mais de 5000 dados sobre iniciativa no âmbito das ODS desde 2020 (ano de implementação) até 2024 e foi ainda efetuada a recolha de mais de 1200 dados sobre preços dos transportes desde 2010 até 2024.

Foi na base de todas estas análises que foi possível obter a informação hoje disponibilizada na Pegada Lusa. A análise sentimental de todos os artigos acima mencionados permitiu gerar uma pontuação a cada distrito, que foi batizada de “*GreenScore*”, e tem uma escala de 0 a 100%. Nesta pontuação, não existiu uma grande discrepância entre distritos, sendo que Lisboa conseguiu a nota mais alta com 64,22% em 1011 artigos, enquanto a Madeira ficou na cauda com 46,59% em 1012 artigos.

Nos ODS, Lisboa também lidera desde 2021 no número de projetos e desde 2023 em boas práticas, enquanto Vila Real fica na cauda com nenhuma boa prática implementada e apenas 6 projetos implementados (menos 1 que em 2023). Relativamente aos objetivos, “Educação e Qualidade” é o mais representado em projetos e boas práticas desde 2020, enquanto “Energias Renováveis” é o menos representado nos projetos desde 2022, e “Vida Debaixo de Água” o menos representado em boas práticas desde 2020.

Nos transportes, verifica-se que os preços das viagens de carro (através das portagens e combustíveis) e de comboio têm tendência a aumentar, já o avião e o autocarro são voláteis, mas com tendência a diminuir consoante a concorrência a efetuar o trajeto. Considerando uma viagem com cinco passageiros e as emissões poluentes de cada transporte associados à viagem, verifica-se que as viagens menos poluentes são aquelas que têm tendência a aumentar, contribuindo assim para um impacto negativo na pegada ecológica.

Originalidade e caráter inovador

A “Pegada Lusa” é o primeiro projeto a reunir e analisar dados sobre políticas e ações sustentáveis em Portugal, transformando-os em estatísticas e classificações sobre as ações realizadas. Além disso, o facto de apresentar dados regionalizados baseando-se em jornais locais é um método original que visa criar uma maior proximidade ao destinatário.

Também relativamente aos ODS, é o único sítio onde é possível observar as iniciativas entretanto descontinuadas, pois no odslocal.pt apenas é possível visualizar as correntes iniciativas.

Impacto social (aplicação e utilidade social)

A exposição das boas práticas e seus ODS, e dos artigos locais e da sua classificação quanto à sustentabilidade pode consciencializar os cidadãos e as organizações para a adoção de práticas mais verdes. A Pegada Lusa pode servir também como uma plataforma de partilha de iniciativas entre comunidades, e de influência nas políticas públicas e empresariais.

Impacto científico (aplicação e utilidade científica)

A “Pegada Lusa” é uma valiosa fonte de dados para pesquisadores e interessados em sustentabilidade, mudanças climáticas e políticas ambientais. Ao fornecer acesso a informações detalhadas, categorizadas e regionalizadas sobre as iniciativas de sustentabilidade, facilita a análise de tendências e a eficácia de diferentes abordagens.

Relevância da utilização do Arquivo.pt

Todos os artigos jornalísticos analisados pela Pegada Lusa são provenientes do Arquivo.pt, não só porque este disponibiliza dados históricos e versionados dos *websites* portugueses, mas também pela disponibilização da API TextSearch que permite a pesquisa em volume de diferentes fontes de dados (jornais digitais).

Além disso, para a elaboração de uma linha cronológica sobre os ODS e para a ferramenta de análise de preços dos transportes, o Arquivo.pt foi essencial para que pudessem ser recolhidos os dados referentes a anos anteriores.

Recursos complementares

- *Portal de dados abertos da Administração Pública*
<https://dados.gov.pt/s/resources/publicacoes-periodicas-portuguesas-jornais-e-revistas-websites-e-historico-de-versoes-no-arquivo-pt/20210731-173642/publicacoes-periodicas-portuguesas-websites-e-historico-de-versoes-no-arquivopt.xlsx> - *Lista de Jornais e Revistas ERC*