# Covid Data Analytics: Repositório de Dados Provenientes de Múltiplas Fontes sobre a Pandemia de COVID-19 no Brasil

**Pedro Moreira[1], Rodrigo Fonseca[1], Pedro Loures Alzamora[1], Ramon A. S. Franco[2], Janaina Guiginski[1], Evandro L. T. P. Cunha[1], Tereza Bernardes[1], Bruno Chagas[1], Kícila Ferreguetti[1], Luana Passos[1], Luísa Cardoso[1], Raquel Schneider[1], Wallace Pereira[1], Ana Paula Couto da Silva[1], Wagner Meira Jr.[1]**

[1]Universidade Federal de Minas Gerais (UFMG)

[2]Universidade Federal do Oeste da Bahia (UFOB)

{pedrovxm,janainaguiginski}@gmail.com, ramon.franco@ufob.edu.br,

{ana.coutosilva,meira}@dcc.ufmg.br

***Abstract.*** *This paper presents the construction and deployment of a data repository used and developed under the Covid Data Analytics (CDA) project, executed by the Department of Computer Science at UFMG. The project aimed to monitor aspects related to the social, economic and epidemiological scenario of COVID-19 in Brazil by analyzing data from official and non official sources, online social networks, and the web in general. The construction of the repository, which contains 18 attributes and 1086 records, was based on collecting data directly from the selected sources, which were later enriched and, finally, made available through a search tool developed exclusively for them.*

***Resumo.*** *Este artigo apresenta a construção e publicação de um repositório de dados utilizados e desenvolvidos no âmbito do projeto Covid Data Analytics (CDA), executado pelo Departamento de Ciência da Computação da UFMG. O projeto visou monitorar aspectos referentes à situação social, econômica e epidemiológica da COVID-19 no Brasil a partir da análise de dados provenientes de fontes oficiais e não oficiais, de redes sociais online e da web em geral. A construção do repositório, contendo 18 atributos e 1086 registros, se baseou na coleta direta de dados das fontes selecionadas, os quais foram posteriormente enriquecidos e, finalmente, disponibilizados por meio de uma ferramenta de busca desenvolvida exclusivamente para eles.*

## 1. Introdução

A pandemia de COVID-19 apresentou grandes desafios para toda a humanidade: além da natureza relativamente desconhecida do patógeno e de suas consequências para a saúde pública, é importante destacar o efeito da adoção de novos hábitos e comportamentos na sociedade, tais como a necessidade de utilização de máscaras e da manutenção de um distanciamento físico-social. O alto índice de contágio foi intensificado por padrões de mobilidade e concentração populacional, bem como por aspectos socioeconômicos, que passaram a ser dimensões necessárias para a compreensão da difusão e da multidimensionalidade do impacto da doença nas populações afetadas.

Uma estratégia para melhor compreender as diversas facetas e possíveis impactos da pandemia de COVID-19 na sociedade consiste na extração de informação e conhecimento a partir de dados provenientes de diversas fontes oficiais e não oficiais. Dados gerados por usuários de redes sociais online também possuem grande potencial de exploração, uma vez que o isolamento físico-social pode ter impulsionado ainda mais a utilização dessas ferramentas. Esse tipo de conteúdo interessa a pesquisadores em diversas áreas do conhecimento, tais como Ciências da Saúde, Economia, Linguística, Política e Demografia, além da própria Ciência da Computação.

A importância desse tema fomentou a publicação de diversos artigos científicos que investigam aspectos relacionados à pandemia de COVID-19 no Brasil por meio de análises de dados. Alguns trabalhos, por exemplo, fornecem caracterizações e descrições da evolução da doença no país [Ranzani et al. 2021], considerando, inclusive, a sub-notificação de casos pelas agências oficiais [Veiga e Silva et al. 2020]. Outros modelam e preveem a evolução da COVID-19, utilizando dados referentes aos primeiros meses da pandemia e empregando diferentes métodos [Bastos and Cajueiro 2020, Pereira et al. 2020], ou mesmo utilizando dados de geolocalização e de dinâmica populacional [Peixoto et al. 2020]. Nesse contexto, é importante que, sempre que possível, os dados utilizados nas pesquisas sejam disponibilizados à comunidade científica e sociedade em geral, seja para fins de replicabilidade dos resultados encontrados, seja para a promoção de novas investigações.

O projeto Covid Data Analytics (CDA)[1], proposto pelo Departamento de Ciência da Computação da Universidade Federal de Minas Gerais (DCC/UFMG) e executado nos anos de 2020 e 2021, integrou uma equipe transdisciplinar, composta por estudantes de graduação e pós-graduação, além de pós-doutorandos e pesquisadores seniores, com a finalidade de melhor compreender os impactos da pandemia na saúde pública, na socioeconomia e no comportamento da população, sobretudo a partir da análise integrada de dados estruturados e de dados coletados da web. Dentre os principais objetivos do projeto CDA, se destaca a divulgação pública de dados em diferentes formatos e provenientes de várias fontes que possam ser utilizados para a investigação dos impactos da pandemia no Brasil. Para tanto, os pesquisadores vinculados ao projeto coletaram, organizaram e analisaram dados epidemiológicos, sociais, econômicos e demográficos. Com a atuação de uma equipe interdisciplinar, proveniente de áreas como a Ciência da Computação, a Demografia, a Economia, a História, a Linguística e a Medicina, foram coletados dados estruturados de fontes oficiais (IBGE[2], PNADs[3], DATASUS[4]) e não oficiais (Brasil.IO[5]), de redes sociais online (Twitter, YouTube e Instagram) e da web em geral (Google Trends).

A equipe do projeto se dividiu em quatro linhas de pesquisa: (i) "Análise do comportamento da economia brasileira"; (ii) "Estratégias de intervenções de telessaúde na pandemia de COVID-19"; (iii) "Indicadores epidemiológicos e comportamento na web"; e (iv) "Política, ideologia e informações médicas nas redes". Cada uma dessas linhas de pesquisa promoveu a coleta de dados de diferentes naturezas, com o objetivo de subsi-

[1]https://covid.dcc.ufmg.br
[2]https://www.ibge.gov.br/
[3]https://www.ibge.gov.br/pnds
[4]http://www2.datasus.gov.br
[5]https://brasil.io/home/

diar as análises a serem realizadas. A grande quantidade e diversidade dos dados coletados fomentou a criação de um grupo exclusivamente responsável pela sua organização e divulgação: a equipe "Coleta e divulgação de dados".

Neste artigo, apresentamos a metodologia utilizada para a construção do repositório *Covid Data Analytics*. Os dados provenientes das coletas realizadas, bem como um rico conjunto de análises realizadas, se encontram disponíveis na plataforma Zenodo[6].

Este artigo está organizado da seguinte forma. A Seção 2 apresenta a metodologia desenvolvida para a construção do repositório de dados. A descrição geral dos dados é apresentada na Seção 3. Exemplos de algumas análises realizadas a partir dos dados coletados são apresentados na Seção 4. A Seção 5 descreve o potencial para futuras análises a partir dos dados do nosso repositório. A Seção 6 conclui o artigo.

## 2. Metodologia para a Construção do Repositório de Dados

A metodologia desenvolvida para a criação do repositório e a integração dos dados coletados e das análises realizadas a partir dos dados coletados é dividida em três etapas principais: (1) identificação das fontes de dados; (2) coleta dos dados; (3) unificação dos dados e disponibilização via portal web. A Figura 1 apresenta a integração destas etapas.

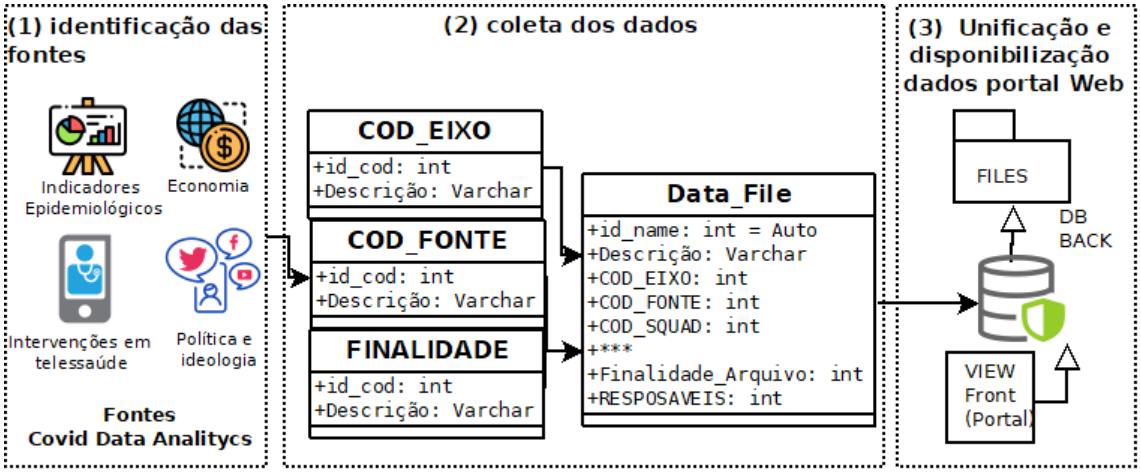


**Figura 1. Metodologia para Construção do Repositório Covid Data Analytics.**

A primeira etapa da metodologia, *identificação das fontes de dados*, teve como objetivo definir as fontes de dados de interesse. Para esta definição, algumas perguntas de pesquisa a serem respondidas foram definidas: *Como se deu a evolução temporal e espacial do novo coronavírus no território brasileiro? Como eventos relacionados à pandemia repercutiram no mundo virtual? Como as tecnologias de telessaúde podem auxiliar no enfrentamento da pandemia da COVID-19? Quais são os principais desafios e oportunidades das intervenções de telessaúde? Como se deu o uso político e ideológico das informações médico-científicas divulgadas nas redes sociais?* A partir dessa lista inicial de perguntas, foram selecionados dados provenientes de diferentes fontes: dados estruturados de fontes oficiais (IBGE, PNADs, DATASUS) e não oficiais (Brasil.IO), dados de redes sociais online (Twitter e YouTube) e dados da web em geral (Google Trends).

[6]https://zenodo.org/record/5176798

A segunda etapa, *coleta dos dados*, compreendeu tanto a coleta dos dados a partir das fontes selecionadas na primeira etapa quanto as respectivas análises a partir dos mesmos, visando a disponibilização via portal Web do projeto. Devido à natureza heterogênea destes dados e análises, foram seguidas estratégias diferenciadas de coleta e organização:

**Dados de fontes oficiais** Os dados socioeconômicos foram obtidos diretamente das fontes institucionais, em arquivos no formato CSV.

**Dados do Brasil.IO (fonte não oficial)** Os dados organizados e consolidados de casos e óbitos foram coletados semanalmente (arquivos no formato CSV).

**Twitter** Os dados foram coletados através da utilização da API oficial do Twitter[7], com coletas semanais entre 23/02/2020 e 08/05/2021.

**Google Trends** Foi usada a *G Trends API Access* para coletas entre 23/02/2020 e 08/05/2021.

**Relatórios e análises** Através de um instrumento de coleta online (Google Forms), foram coletados os resultados das análises desenvolvidas pelas linhas de pesquisa a partir dos dados coletados das fontes selecionadas.

A última etapa da metodologia organizou e unificou os arquivos (bases de dados, tabelas, gráficos, mapas, relatórios, artigos, etc.) para posterior publicação no portal Web. A seguir, detalhamos as subetapas que compõem esta última etapa de unificação dos dados para disponibilização no portal Web:

1. **Enriquecimento dos dados** Nesta subetapa, os arquivos são enriquecidos com os seguintes metadados: *Data, Nome, Descrição, Última atualização, Localização geográfica, Estado atual, Linha de pesquisa, Fonte (origem) dos dados*. Para os dados estruturados, foram também considerados os seguintes metadados: *Técnica de extração, Data e hora da extração, Finalidade do arquivo, Linha de pesquisa responsável*. A descrição dos metadados selecionados segue o padrão *Open Data Standards*[8], divulgado pelo *World Wide Web Consortium* (W3C)[9]
2. **Implementação do banco de dados** Inicialmente, o banco de dados foi criado em uma máquina local, com o auxílio da ferramenta MySQL Workbench e de um código Python para inserção dos arquivos e seus metadados. Em seguida, foi realizada a integração com o Banco de Dados do projeto CDA, hospedado nos servidores do DCC/UFMG. O banco de dados para unificação dos arquivos foi criado com o auxílio do software phpMyAdmin[10].
3. **Implementação da interface de busca** A interface de busca[11] (Figura 2) permite o acesso aos arquivos e seus metadados por meio de consultas orientadas por termo de busca ou filtros (tipo, fonte ou tema).
   Através desta interface, diferentes tipos de busca podem ser realizadas. Por exemplo, o campo pode conter o termo de busca *número de casos* e selecionar o filtro Tipo, indicando o tipo de arquivo (mapa, gráfico, texto, etc.). Alternativamente, pode-se habilitar o filtro por Fonte, que agrupa os documentos segundo a origem dos mesmos (por exemplo, *YouTube, Google ou IBGE*, ou por Tema. A Figura 3

[7] https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets
[8] https://standards.theodi.org/
[9] https://www.w3.org/
[10] https://www.phpmyadmin.net/
[11] https://covid.dcc.ufmg.br/buscador.php


**Figura 2. Interface de busca do portal Web do projeto CDA.**

exibe o resultado ao buscar pelo termo ”Covid-19”. A interface disponibiliza as informações mais relevantes sobre o arquivo e fornece a URL para download dos dados e informações no repositório CDA.

Como mencionado, os arquivos com todos os materiais coletados e desenvolvidos ao longo do projeto - bases de dados, tabelas, gráficos, mapas, relatórios, artigos, etc. - também estão disponíveis publicamente na plataforma Zenodo.[12]

## 3. Descrição do Conjunto de Dados

Os dados disponibilizados no repositório CDA se referem ao período entre 23 de fevereiro de 2020 e 8 de maio de 2021. Esse repositório agrega 1.086 arquivos, classificados em dois tipos principais: (i) bases de dados e tabelas extraídas das fontes descritas anteriormente; e (ii) artigos, relatórios, mapas e gráficos produzidos pelos integrantes do projeto a partir da análise dos dados coletados [Moreira et al. 2021].

### 3.1. Dados de Fontes Externas

Estes arquivos representam 8% do total de arquivos que compõem o repositório e estão distribuídos da seguinte maneira:

- 71 séries temporais com indicadores econômicos das Unidades Federativas do Brasil e da União em formato .csv, com aproximadamente 18.400 registros;
- 7 scripts de tratamento de dados em formato .py.
- 5 arquivos com a contagem do número de tweets e retweets coletados semanalmente utilizando 13 palavras-chave (“corona”, “covid”, “coronavirus”, “covid19”, “quarentena”, “hidroxicloroquina”, “cloroquina”, “confinamento”, “distanciamento social”, “aglomeração”, “aglomerações”, “sars” e “covid-19”), no formato .csv
- 3 arquivos do Google Trends no formato .csv com 249 registros contendo 124 termos pré-selecionados que têm relação com a pandemia e o percentual relativo de buscas na web nos níveis regional e nacional.

[12]https://doi.org/10.5281/zenodo.5176798

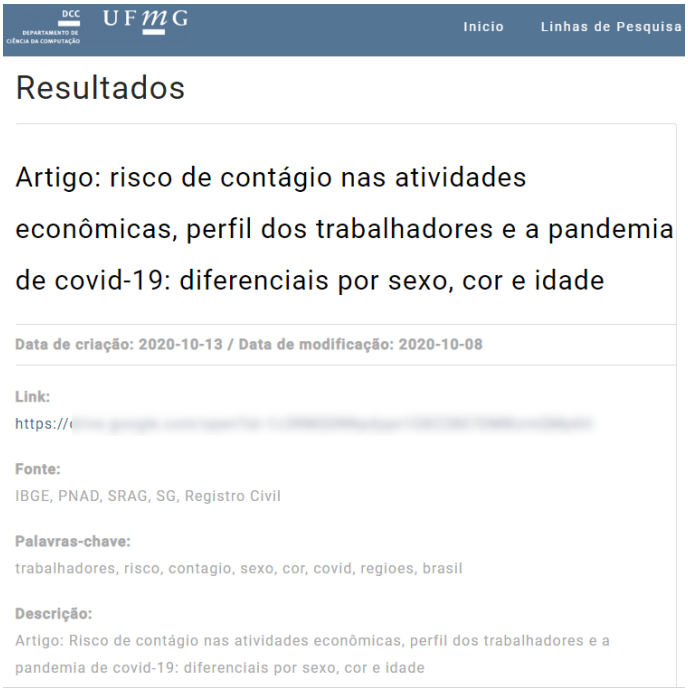


**Figura 3. Resultado da busca pelo termo ”Covid-19”.**

### 3.2. Análises e Relatórios

Os arquivos com as análises e relatórios representam 92% do total de arquivos do repositório. Além de documentos de texto, também foram disponibilizados materiais visuais, como mapas e gráficos, em diversos formatos. Os arquivos estão distribuídos da seguinte maneira:

- 23 gráficos comparativos de indicadores sociais e econômicos, análises descritivas dos ocupados em atividades essenciais e não essenciais por região em formato .svg;
- 409 gráficos de novos casos e óbitos em formato .png;
- 522 mapas e gráficos de linhas e barras acerca dos casos e óbitos acumulados de COVID-19 em todo o país entre as semanas epidemiológicas 9 e 32 de 2020 (23/02/2020 a 08/08/2020) em formato .png;
- 15 arquivos de Medidas Provisórias em formato .pdf;
- 1 gráfico interativo gerado a partir do cálculo da mortalidade (óbitos acumulados por 100 mil habitantes) no Brasil em formato .html;
- 25 animações mostrando a evolução da letalidade (mortes acumuladas / casos acumulados) em % em todos estados do Brasil a cada semana epidemiológica da 9ª à 31ª em formato .gif;
- 1 relatório sobre análises das informações disponíveis para coleta na ferramenta Google Trends em formato .pdf;
- 4 relatórios sobre análises das informações disponibilizadas pelas linhas de pesquisa, em formato .pdf;

### 3.3. Limitações nas Bases de Dados Disponibilizadas

Devido a questões de privacidade e proteção de dados, algumas bases de dados obtidas por meio da extração de informações de redes sociais online não foram integralmente

disponibilizadas no repositório. Nesses casos, foram disponibilizadas análises realizadas utilizando essas bases, executadas com o propósito de responder a algumas das perguntas de pesquisa do projeto. Informações sobre as análises realizadas durante a vigência do projeto estão disponíveis em `https://covid.dcc.ufmg.br/`.

## 4. Exemplos de Análises

A seguir são apresentados, de forma breve, dois exemplos de utilização dos dados, disponíveis no repositório CDA. O primeiro exemplo consiste em um estudo da evolução temporal e espacial da COVID-19 no Brasil. O segundo exemplo apresenta um estudo de análise léxica das postagens no Twitter e sua correlação com indicadores epidemiológicos da COVID-19. Análises complementares e de outros trabalhos podem ser encontrados no portal oficial do projeto CDA.[13]

### 4.1. Análise temporal e espacial dos casos de COVID-19 no Brasil

Para realizar esta análise, utilizamos dados de casos e óbitos de COVID-19, extraídos originalmente do Brasil.IO, por dia e por município. Após a extração, os dados foram pré-processados, de forma a agregar as informações em semanas e regiões imediatas. O desenvolvimento das análises espaciais baseou-se em técnicas de Análises Exploratórias de Dados Espaciais (AEDE) [Rey S. J. 2020]. O geoprocessamento de dados e as análises espaciais são ferramentas importantes para o estudo de fenômenos como a disseminação de doenças, no caso da pandemia, que se espalha pelo território brasileiro de forma heterogênea no tempo e no espaço.

Para representar os padrões espaciais de distribuição de casos e óbitos de COVID-19, foram gerados mapas, que representam dois períodos da pandemia: o primeiro, que se estende de abril a agosto de 2020 e o segundo, que abrange novembro de 2020 a março de 2021. Os diversos mapas com os indicadores epidemiológicos analisados, como o total de óbitos e total de casos por 100.000 habitantes, e a letalidade da doença, estão disponíveis no repositório. Como exemplo dessas análises, a Figura 4 apresenta um dos mapas desenvolvidos durante a análise espacial: a distribuição dos casos semanais de COVID-19, que destaca a formação de agrupamentos (*clusters*) de regiões com alto número de casos, representados na cor vermelha, e regiões com menor quantidade relativa de casos da doença, representadas em azul.

### 4.2. Análise lexical das postagens no Twitter e sua correlação com indicadores epidemiológicos

A utilização da análise de redes sociais online como ferramenta para a compreensão de fenômenos relacionados à Saúde Pública é uma prática amplamente adotada no meio científico. O monitoramento de redes sociais online permite acompanhar, prever e avaliar a repercussão destes fenômenos em tempo real [Du et al. 2017, Sultana et al. 2021, Kang et al. 2017]. Neste contexto, o Twitter é uma rede comumente utilizada [Marques-Toledo et al. 2017, Gomide et al. 2011, Aiello et al. 2020, Li et al. 2020], uma vez que sua política de coleta de dados permite obter grandes quantidades de dados, filtrados por palavras-chave relacionadas ao tema de interesse. Além disso, trata-se de uma rede social de caráter extremamente reativo, o que favorece estudos em tempo real.

[13] https://covid.dcc.ufmg.br

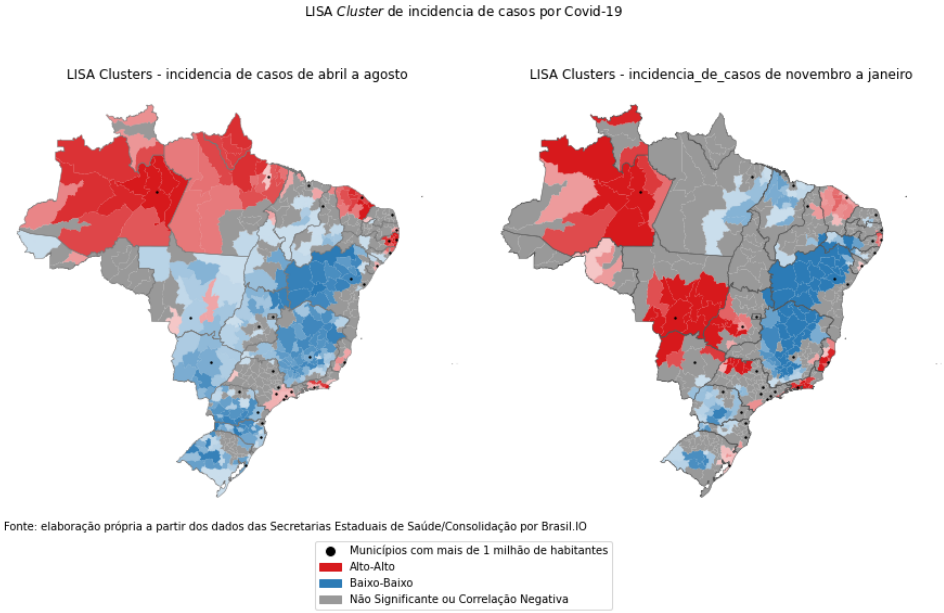


**Figura 4. Mapas de agrupamentos de novos casos de COVID-19 por 100.000 habitantes**

Nesta análise, foram utilizados dados coletados do Twitter e dois indicadores epidemiológicos - os fatores de crescimento de casos e de mortes de COVID-19 para o Brasil. Tabelas com estes indicadores, entre outros, estão disponíveis no repositório. Também constam no repositório tabelas com a contagem de palavras presentes em publicações do Twitter relacionadas à pandemia, no período entre março de 2020 e janeiro de 2021.

A partir desses dois conjuntos de dados (indicadores e *tweets*), foi possível identificar os temas de discussão no Twitter que estavam mais frequentemente correlacionados com o aumento ou com a redução dos casos e óbitos. Para exemplificar, ao considerarmos discussões realizadas duas semanas antes do aumento no número de óbitos, as palavras mais frequentes encontradas nos *tweets* revelam uma mistura de sentimentos: *rir, piada, alegre, testou, pico*.

## 5. Potencial do Repositório CDA

O potencial de contribuição do repositório de dados do projeto CDA reside na integração dos dados de um projeto multidisciplinar. Esta integração procura superar o obstáculo acarretado pela descentralização e diversidade das fontes, permitindo a consulta dos dados, artefatos e análises por qualquer pesquisador (ou cidadão em geral) interessado no tema da pandemia de COVID-19 no Brasil. Futuramente, espera-se que a disponibilização dos arquivos ao público tenha o potencial de estender as fronteiras de colaboração internacional em pesquisas que visam compreender a pandemia a partir de uma perspectiva regional.

A metodologia implementada permitiu enriquecer mais de 1000 arquivos com diferentes formatos e aplicações. Os arquivos foram disponibilizados por equipes compostas por especialistas das áreas de Economia, Linguística, Computação, Medicina, História, Demografia, Ciências Humanas e Sociais. Como resultado, temos uma grande variedade de análises interdisciplinares da pandemia de COVID-19 no Brasil, desde o seu

início, em 2020, até meados de maio de 2021.

Concentramos nossos esforços na divulgação das informações, pois acreditamos que os arquivos disponibilizados poderão ser utilizados em outros estudos. Dada a complexidade e gravidade das consequências da pandemia em todo o território nacional, acreditamos que armazenar e disponibilizar informações em um repositório unificado tem o potencial de auxiliar o entendimento da situação atual e o planejamento de ações futuras. Assim, esperamos que a implementação de um buscador amigável contribua para a ampla consulta e utilização dos dados, artefatos e análises, beneficiando novas pesquisas interdisciplinares, tanto pelos pesquisadores que participaram do projeto quanto por pesquisadores externos.

## 6. Considerações Finais

Neste artigo é apresentada a construção e publicação de um repositório que agrega e sistematiza dados provenientes de fontes oficiais (IBGE, PNADs, DATASUS), fontes não oficiais (Brasil.IO), redes sociais online (Twitter, YouTube) e web (Google Trends) no contexto da pandemia de COVID-19 no Brasil. Apresenta-se, também, uma visão geral acerca do amplo conjunto de análises executadas pelos diferentes grupos de pesquisa vinculados ao projeto Covid Data Analytics (CDA), do Departamento de Ciência da Computação da UFMG, as quais são também parte do repositório.

Acreditamos que estes dados e resultados podem ser utilizados em várias análises sobre o impacto da COVID-19 no Brasil, uma vez que a pandemia ainda segue em curso e ainda existe muita incerteza sobre os seus desdobramentos.

**Agradecimentos.** Este trabalho foi realizado com apoio financeiro do CNPq, FAPEMIG, CAPES e dos projetos Covid Data Analytics (PRPq/UFMG/SESU/MEC), MASWEB, INCT-Cyber e Atmosphere.

## Referências

Aiello, A. E., Renson, A., and Zivich, P. N. (2020). Social media– and internet-based disease surveillance for public health. *Annual Review of Public Health*, 41(1):101–118. PMID: 31905322.

Bastos, S. B. and Cajueiro, D. O. (2020). Modeling and forecasting the early evolution of the Covid-19 pandemic in Brazil. *Scientific Reports*, 10(1):1–10.

Du, J., Xu, J., Song, H., Liu, X., and Tao, C. (2017). Optimization on machine learning based approaches for sentiment analysis on hpv vaccines related tweets. *Journal of biomedical semantics*, 8(1):1–7.

Gomide, J., Veloso, A., Meira, W., Almeida, V., Benevenuto, F., Ferraz, F., and Teixeira, M. (2011). Dengue surveillance based on a computational model of spatio-temporal locality of twitter. In *Proceedings of the 3rd International Web Science Conference*, WebSci '11, New York, NY, USA. Association for Computing Machinery.

Kang, G. J., Ewing-Nelson, S. R., Mackey, L., Schlitt, J. T., Marathe, A., Abbas, K. M., and Swarup, S. (2017). Semantic network analysis of vaccine sentiment in online social media. *Vaccine*, 35(29):3621–3638.

Li, C., Chen, L. J., Chen, X., Zhang, M., Pang, C. P., and Chen, H. (2020). Retrospective analysis of the possibility of predicting the covid-19 outbreak from internet searches and social media data, china, 2020. *Eurosurveillance*, 25(10).

Marques-Toledo, C. d. A., Degener, C. M., Vinhal, L., Coelho, G., Meira, W., Codeço, C. T., and Teixeira, M. M. (2017). Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting dengue at country and city level. *PLoS neglected tropical diseases*, 11(7):e0005729.

Moreira, P. V. X., Franco, R. A. S., Fonseca, R. M., Prado, A. C. T., Leal, L., Mendes, G. N., and Rezende, T. A. V. (2021). Covid Data Analytics: Repositório de Dados Provenientes de Múltiplas Fontes sobre a Pandemia de COVID-19 no Brasil https://doi.org/10.5281/zenodo.5176798. *Zenodo*.

Peixoto, P. S., Marcondes, D., Peixoto, C., and Oliva, S. M. (2020). Modeling future spread of infections via mobile geolocation data and population dynamics. an application to COVID-19 in Brazil. *PloS one*, 15(7):e0235732.

Pereira, I. G., Guerin, J. M., Silva Júnior, A. G., Garcia, G. S., Piscitelli, P., Miani, A., Distante, C., and Gonçalves, L. M. G. (2020). Forecasting Covid-19 dynamics in Brazil: a data driven approach. *International Journal of Environmental Research and Public Health*, 17(14):5115.

Ranzani, O. T., Bastos, L. S., Gelli, J. G. M., Marchesi, J. F., Baião, F., Hamacher, S., and Bozza, F. A. (2021). Characterisation of the first 250 000 hospital admissions for COVID-19 in Brazil: a retrospective analysis of nationwide data. *The Lancet Respiratory Medicine*, 9(4):407–418.

Rey S. J., Arribas-Bel D., W. L. J. (2020). Geographic data science with pysal and the pydata stack.

Sultana, A., Tasnim, S., Hossain, M. M., Bhattacharya, S., and Purohit, N. (2021). Digital screen time during the covid-19 pandemic: a public health concern. *F1000Research*, 10(81):81.

Veiga e Silva, L., de Andrade Abi Harb, M. D. P., Dos Santos, A. M. T. B., de Mattos Teixeira, C. A., Gomes, V. H. M., Cardoso, E. H. S., da Silva, M. S., Vijaykumar, N., Carvalho, S. V., Frances, C. R. L., et al. (2020). COVID-19 mortality underreporting in Brazil: analysis of data from government internet portals. *Journal of medical Internet research*, 22(8):e21413.