

# Manual do usuário

# PDD

09/2024

—

Plataforma de Dados Desidentificados

Cidacs/Fiocruz - BA

## SUMÁRIO

1. Apresentação.....	3
I. Introdução .....	3
II. Estruturação e desenvolvimento da plataforma .....	4
III. Principais características e vantagens .....	5
IV. Resultados e potencialidades.....	6
2. Aplicação Web.....	9
I. Tela de Cadastro .....	9
II. Tela de Login .....	10
III. Tela Inicial.....	10
1) Que tipo de dados você vai encontrar aqui? .....	11
2) Catálogo de metadados .....	11
3) Formato dos dados .....	11
4) Visualização de dados .....	11
5) Como ter acesso a estes dados? .....	11
IV. A barra de navegação .....	12
V. Datasets .....	12
VI. Adicionar Datasets ao Perfil .....	13
VII. Meus Datasets .....	14
VIII. Sobre o APP Desktop .....	14
IX. Editar Perfil.....	15
3. Aplicação Desktop.....	16
4. API da PDD.....	19
5. Metadados (Dataverse) .....	22
I. A plataforma de Governança de Dados .....	22
a. Dataverse .....	22
b. A nossa Governança de Dados .....	22
II. O Dataverse do CIDACS .....	23
a. Busca.....	25
b. Informações de um dataset.....	26
c. Citação dos dados.....	30

# 1. Apresentação

Este manual visa apresentar o funcionamento da Plataforma de Dados Desidentificados (PDD) e sua potencialidade de uso na perspectiva de desenvolvimento no Centro de Integração de Dados e Conhecimentos para Saúde (Cidacs) / Fiocruz Bahia. Esta iniciativa surgiu no início da pandemia de Covid-19 em 2020 mediante a dificuldade de obtenção de dados de forma ágil e oportuna para produzir conhecimentos úteis para o monitoramento e avaliação de emergências sanitárias com intuito de apoiar a tomada de decisão por gestores. No âmbito da concepção e desenvolvimento desta plataforma, foi priorizado o levantamento e a disponibilização do acesso de forma centralizada aos dados sobre saúde disponíveis em domínio público, abertos e desidentificados, bem como à integração dessas informações e à produção de conhecimento através de estudos e análises. Ao longo do tempo, foram se integrando outras áreas de conhecimento ligadas a áreas socioeconômicas e mais recentemente dados climatológicos.

## I. Introdução

A Plataforma de Dados Desidentificados (PDD) é uma solução do Centro de Integração de Dados e Conhecimentos para Saúde (Cidacs) voltada para a **disponibilização do acesso a um ecossistema de dados** cujo objetivo é subsidiar a produção de conhecimento em saúde no Brasil. A PDD destina-se ao mapeamento, obtenção, organização, processamento e disponibilização de informações relevantes com potencial de apoiar pesquisadores e gestores na vigilância epidemiológica, monitoramento das emergências de saúde pública, análise da situação de saúde, construção/atualização de indicadores e painéis, pesquisa em determinantes sociais, desigualdades em saúde, avaliação de impacto, etc. O escopo da plataforma engloba os dados de saúde administrativos, abertos, desidentificados (não-nominais), individualizados ou agregados provenientes de fontes de domínio público.

A criação da Plataforma de Dados Desidentificados (PDD) surgiu em resposta à necessidade premente de aprimorar a eficiência na obtenção de dados de saúde advindos dos diversos e heterogêneos sistemas de informações e fontes de dados no Brasil, sobretudo durante eventos críticos como foi a pandemia de Covid-19. Nesse contexto, a plataforma foi concebida reconhecendo a importância da centralização de informações ágeis, úteis e de qualidade em um único ambiente como estratégia para

aprimorar a capacidade de análise e pesquisa, contribuir no entendimento e enfrentamento de crises sanitárias e apoiar a vigilância epidemiológica em saúde.

## II. Estruturação e desenvolvimento da plataforma

A plataforma possui uma governança de dados composta pela coordenação, captação, tratamento, gestão do acesso e análise dos dados. A camada de captação de dados é responsável pelo mapeamento de fontes de dados e informações úteis, obtenção dos dados (*download*), pré-processamento, ingestão no repositório, catalogação e geração de metadados. Neste nível são aplicados processos automatizados de ETL de dados. A camada de tratamento inclui a preparação dos dados no repositório, isto é, a configuração para permitir a consulta e extração pelos usuários, o enriquecimento, a padronização, a harmonização e a construção de outros datasets a partir dos dados originais (ex. dados em painel). O nível seguinte diz respeito ao acesso dos dados e metadados disponibilizados pela plataforma. Atualmente, é composto por três interfaces: o site da plataforma que gerencia o acesso dos usuários ao catálogo de dados; o Dataverse Cidacs, ferramenta que abriga a coleção de metadados dos datasets da plataforma; e a extração de dados caracterizada pelas ferramentas desenvolvidas para o consumo dos dados da plataforma (aplicativo *desktop*, API's e biblioteca). Por último, tem-se a camada de análise onde estudos são desenhados e executados a partir dos dados disponíveis no ecossistema da PDD.

A arquitetura da plataforma possui uma infraestrutura em nuvem baseada na Google Cloud Platform (GCP) que pode ser compreendida em cinco níveis: dados, *data warehouse*, metadados, dados administrativos e usuários. Para operacionalizar a coleta de dados provenientes de diversas fontes e, também, pensando na garantia da atualização desses dados, a plataforma implementa um processo ETL (extract, transform e load) automatizado através de um ambiente de captação que executa *scripts* em linguagem de programação Python por meio de uma ferramenta que gerencia as rotinas automáticas, o Apache Airflow. O BigQuery, serviço de *data warehouse* (armazenamento de dados) da Google, é utilizado como repositório por ser uma solução altamente escalável e permitir a realização de consultas rápidas usando linguagem SQL mesmo lidando com um alto volume de dados. Essa escolha tecnológica viabiliza não apenas o armazenamento eficiente, mas também o tratamento lógico das tabelas, a produção de novos *datasets* e o desenvolvimento de aplicações para o consumo dos dados, aproveitando todas as ferramentas de conexão com *datalake* e exploração dos

dados fornecidas pela própria *Cloud*. Os metadados coletados e gerados sobre os *datasets* da plataforma (dicionário de dados, notas metodológicas, artigos, etc) são armazenados na coleção da PDD dentro do Dataverse, ferramenta institucional do Cidacs destinada à publicação de metadados, onde recebem um identificador persistente (DOI) para garantir a integridade e rastreabilidade. As informações administrativas utilizadas para o gerenciamento dos usuários e acesso aos dados na plataforma são armazenadas em um banco de dados MySQL. Este banco de dados é consumido pelas interfaces construídas para os usuários: o site da PDD, que é uma aplicação *Web* criada em *Ruby on Rails*; o aplicativo desktop que possui uma interface gráfica para extração dos dados, desenvolvida em Python e executável em Windows, Linux e MacOS; a biblioteca é um cliente Python implementada para consumo dos dados via API diretamente em ambientes de programação.

### III. Principais características e vantagens

A Plataforma de Dados Desidentificados (PDD) apresenta algumas vantagens que a tornam uma ferramenta robusta e flexível para análise em saúde. Além da diversidade de dados que abrange, a arquitetura computacional adotada garante a escalabilidade do armazenamento e processamento, a automação do ETL e a variedade de ferramentas disponíveis e possíveis de serem implementadas para o consumo e acesso aos dados. Isto contribui para criar uma plataforma ágil e adaptável, atendendo às crescentes complexidades das análises em saúde.

- a. **Governança estruturada:** a PDD possui uma governança de dados estruturada, abrangendo coordenação, captação, tratamento, acesso e análise. Essa abordagem organizada e abrangente permite uma gestão eficiente e segura dos dados, garantindo a confiabilidade, proteção e integridade das informações disponibilizadas.
- b. **Escopo dinâmico e maior abrangência de análise:** em comparação com outras iniciativas, a PDD destaca-se pelo escopo dinâmico de dados, oferecendo maior abrangência de análise e possibilidade de atender um público-alvo diversificado. Além de dados específicos de saúde, a plataforma inclui conjuntos diversos, como dados econômicos, demográficos, censitários, amostra por domicílio (PNAD), mobilidade, indicadores, programas sociais, proporcionando uma visão mais holística e completa.

- c. **Extração de dados:** a plataforma permite uma extração de dados voltada para o atendimento das demandas específicas dos usuários através de interfaces interativas que enriquecem e facilitam a experiência, à medida que possibilitam a seleção de conjuntos de dados, variáveis, períodos e áreas geográficas e permitem a visualização prévia da estrutura dos dados e da consulta.
- d. **Acesso controlado e aberto com autenticação:** a PDD adota um modelo de acesso aberto com autenticação, equilibrando a disponibilidade dos dados e do custo dos recursos computacionais-financeiros com a segurança e controle de acesso. Essa abordagem permite atender a uma ampla gama de usuários, incluindo gestores, profissionais de saúde, pesquisadores e estudantes.

Um aspecto fundamental da plataforma é que, desde sua definição, focamos em sua capacidade de gerir e centralizar dados provenientes de diversas fontes. Nosso objetivo sempre foi facilitar o acesso a esses dados, então as alterações realizadas nos conjuntos de dados são mínimas, destinando-se apenas à adequação à ferramenta utilizada. Assim, efetuamos pouco ou nenhum tratamento adicional na disponibilização dos dados.

## IV. Resultados e potencialidades

Os dados administrativos públicos de saúde provêm de uma vasta gama de fontes, refletindo a complexidade do sistema de saúde no Brasil e seus inúmeros atores, à exemplo do DATASUS, secretarias estaduais e municipais de saúde, Controladoria Geral da União - CGU (Portal da Transparência), instituições de saúde, portais do Ministério da Saúde, IBGE, etc. Essas fontes de dados consistem em um conjunto heterogêneo que inclui registros hospitalares, relatórios epidemiológicos, bases de dados de seguridade social, mortalidade, nascimentos, registros de notificações de doenças, entre tantos outros. Cada fonte traz consigo nuances específicas, desde diferentes formas de coleta até variações nas definições e padrões de codificação. Além de variar quanto à abrangência geográfica e de período, também divergem quanto à natureza, podendo ser agregada, fornecendo uma visão consolidada e resumida utilizando geralmente estatísticas populacionais ou resultados de análises epidemiológicas em nível macro, ou ainda do tipo de microdados, oferecendo uma perspectiva mais detalhada, consistindo em informações individuais que podem ser utilizadas para análises mais específicas e personalizadas.

Portanto, para sumarizar o ecossistema de dados, as bases de dados foram divididas em treze sessões temáticas, descritas segundo suas características: notificações de casos e óbitos por Covid-19; medidas de prevenção e controle da Covid-19; vacinas; nascimentos e mortes; hospitalizações; assistência e serviços de saúde; dados demográficos; dados socioeconômicos; programas sociais; doenças de notificação compulsória; dados nutricionais; dados ambientais e climáticos; e dados geoespaciais.

**Quadro 1.** Ecossistema de dados da Plataforma de Dados Desidentificados (PDD-Cidacs)

TEMA	DESCRIÇÃO	PRINCIPAIS FONTES DE DADOS
Notificações de casos e óbitos por Covid-19	Dados agregados e individualizados sobre os casos e óbitos por Covid-19 no Brasil a partir de 2020	Ministério da Saúde; e Secretarias Estaduais e Municipais de Saúde
Medidas de prevenção e controle da Covid-19	Medidas adotadas a partir do início da pandemia de Covid-19, incluindo dados sobre mobilidade em tempos de distanciamento social, legislação e o programa federal “Auxílio Emergencial”	Google; Plataformas Leis Municipais e Leis Estaduais; Portal da Transparência da Controladoria Geral da União
Vacinação contra Covid-19	Doses aplicadas de imunização contra Covid-19	Ministério da Saúde, Sistema de Informação do Programa Nacional de Imunizações (SI – PNI)
Nascimentos e mortes	Dados individualizados sobre nascidos vivos e óbitos por todas as causas	Ministério da Saúde, Sistema de Informação sobre Nascidos Vivos (SINASC) e Sistema de Informação sobre Mortalidade (SIM)
Hospitalizações	Dados individualizados de hospitalizações por todas as causas (incluindo Síndrome Respiratória Aguda Grave – SRAG).	Ministério da Saúde, Sistema de Informação Hospitalares do SUS (SIH-SUS) e Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe)
Assistência e serviços de saúde	Conjunto de informações acerca de estabelecimentos de saúde, equipamentos, profissionais, leitos, atenção básica, etc.	Ministério da Saúde, Cadastro Nacional de Estabelecimentos de Saúde - CNES
Dados demográficos e socioeconômicos	Informações e indicadores demográficos/socioeconômicos para estudos em saúde (incluindo dados	Instituto Brasileiro de Geografia e Estatística - IBGE; Governo Federal,



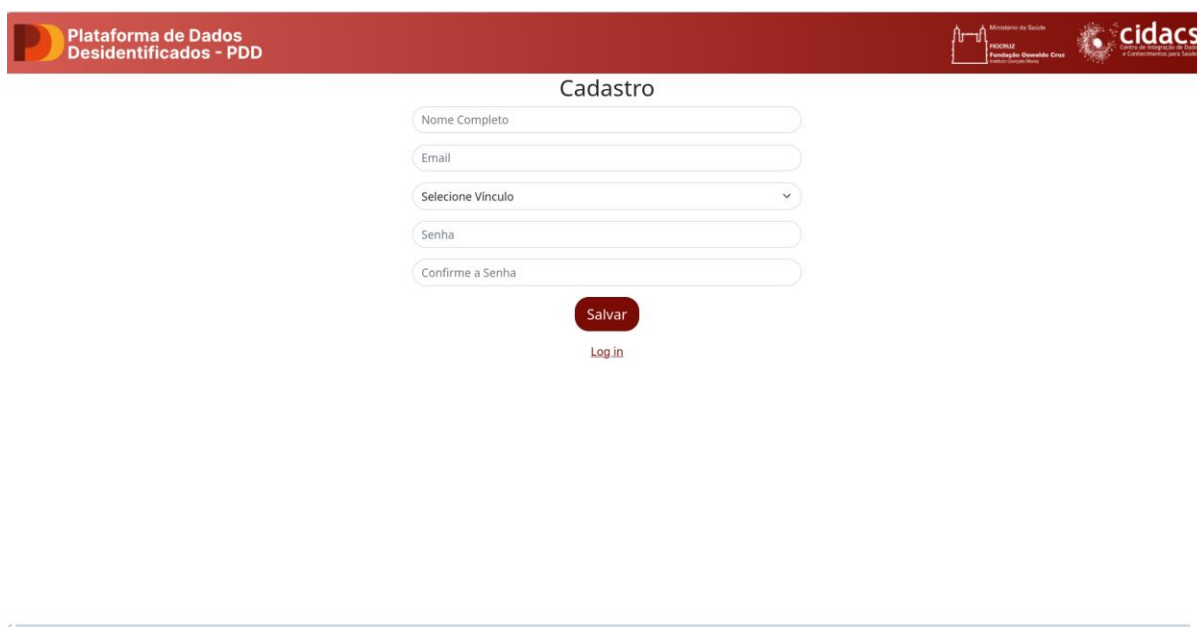
	censitários.	Ministério da Saúde; etc.
Programas sociais	Dados sobre programas sociais federais para estudos de avaliação de impacto (Ex.: Programa Bolsa Família).	Governo Federal, Ministério da Cidadania.
Doenças de notificação compulsória	Dados individualizados sobre doenças e agravos de notificação compulsória oriundas do Sistema de Informação de Agravos de Notificação - SINAN.	Ministério da Saúde, Sistema de Informação de Agravos de Notificação - SINAN.
Dados nutricionais	Dados da Vigilância Alimentar e Nutricional da Atenção Básica	Ministério da Saúde; Sistema de Vigilância Alimentar e Nutricional – SISVAN.
Dados ambientais e climáticos	Dados sobre ambiente e clima, com foco na descrição das características ambientais dos municípios brasileiros e informações úteis para o entendimento das mudanças climáticas.	Ministério da Ciência, Tecnologia e Inovações, Instituto Nacional de Pesquisas Espaciais – INPE; e outras plataformas de pesquisa e disseminação de informações ambientais e climáticas.
Educação	Dados e indicadores educacionais (ex.: IDEB).	Ministério da Educação.
Dados geoespaciais (shapefiles)	Dados auxiliares que representam divisões administrativas e recortes do território brasileiro (ex.: <i>shapefiles</i> dos municípios, unidades da federação, regiões de saúde, etc.	Instituto Brasileiro de Geografia e Estatística - IBGE; Ministério da Saúde; etc;



## 2. Aplicação Web

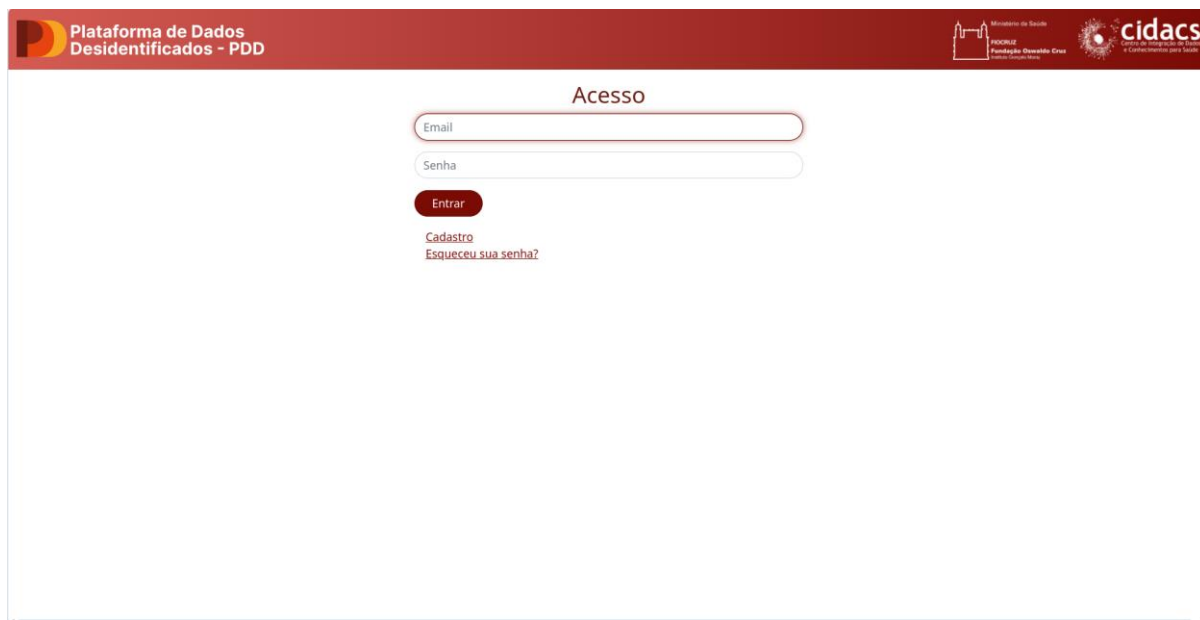
Para acesso aos dados pela Plataforma, os usuários devem visitar o *Web Site* (<https://pdd.cidacs.org/>) e realizar o seu cadastro. Para o cadastro pedimos apenas nome completo, um e-mail para acesso, um vínculo para entendermos o perfil dos nossos usuários e uma senha. Caso o usuário já seja cadastrado, é possível apenas realizar o login criado.

### I. Tela de Cadastro



The screenshot shows the registration page of the PDD (Plataforma de Dados Desidentificados) web application. The page has a dark red header with the PDD logo on the left and the logos of the Ministério da Saúde, FIOCRUZ, and cidacs on the right. The main content area is white and titled "Cadastro". It contains a form with five input fields: "Nome Completo", "Email", "Selecione Vínculo" (a dropdown menu), "Senha", and "Confirme a Senha". Below the form is a red "Salvar" button and a "Log in" link. The page is framed by a light blue border at the bottom.

## II. Tela de Login



The screenshot shows the login page of the Plataforma de Dados Desidentificados (PDD). The header includes the PDD logo and the Cidacs logo. The main content area is titled 'Acesso' and contains a login form with fields for 'Email' and 'Senha' (Password). Below the password field is an 'Entrar' (Login) button. There are also links for 'Cadastro' (Registration) and 'Esqueceu sua senha?' (Forgot your password?).

Após o cadastro ou login os usuários serão redirecionados para a tela inicial onde é apresentada a plataforma e um resumo rápido do que pode ser encontrado e como utilizá-la.

## III. Tela Inicial



The screenshot shows the home page of the Plataforma de Dados Desidentificados (PDD). The header includes the PDD logo, the user name 'Mikael Marin Coletto', and the Cidacs logo. The main content area is titled 'Bem-vindo, Mikael Marin Coletto!' and contains a welcome message, a list of links, and a section titled 'Formato dos dados' (Data Format). The 'Formato dos dados' section explains that data can be downloaded in CSV format and provides a step-by-step guide for obtaining data.

**Bem-vindo, Mikael Marin Coletto!**

A Plataforma de Dados Desidentificados (PDD) visa disponibilizar um ecossistema de dados para a produção de conhecimento em saúde, em especial sobre a pandemia de Covid-19, usando dados administrativos desidentificados individualizados e agregados. Para tal, identifica, obtém, organiza, processa e disponibiliza dados e informações que sejam úteis para realizar as análises pelos pesquisadores.

**Que tipo de dados você vai encontrar aqui?**

Você poderá acessar dados administrativos de domínio público, abertos, não-nominais, agregados e microdados. São dados sobre informações demográficas e socioeconômicas, nutricionais, hospitalizações, nascimentos, óbitos, assistência e serviços de saúde, dados geoespaciais, além de casos, óbitos, testes diagnósticos e exames laboratoriais para Covid-19, vacinação contra Covid-19, medidas de prevenção e controle como o distanciamento social entre outros.

**Catálogo de metadados**

Para cada conjunto de dados disponível na plataforma, existe um conjunto de características como fonte de informação, dicionário de variáveis, data de publicação, entre outros, ou seja, cada conjunto de dados possui seus próprios metadados, para melhor compreensão dos conteúdos.

Os metadados estão disponíveis no [Dataverse](#), que é um repositório de dados de pesquisas gratuito e de código-aberto, de fácil navegabilidade, que permite melhor gestão dos dados e metadados.

**Como ter acesso a estes dados?**

É simples. Para ter acesso aos dados, primeiro adicione os datasets de interesse

**Formato dos dados**

Os dados podem ser baixados em formato CSV (Comma-separated values)

**Passo-a-passo para obter os dados:**

1. Uma vez cadastrado e logado no portal Web da PDD, você poderá:
  - 1.1) Acessar ao catálogo dos datasets disponíveis na plataforma. Ao clicar no botão "Datasets" na barra de menu superior, você obterá o conjunto de dados disponíveis. Ao clicar em cada dataset, você terá a descrição da base e uma amostra da estrutura dos dados;
  - 1.2) Acessar o catálogo de metadados no Dataverse para obter a descrição, os dicionários e os identificadores persistentes das bases de dados (menu "Datasets", barra de menu superior - selecione o dataset - botão "Metadados");
  - 1.3) Explorar os dados através da visualização de dados (menu "Datasets", barra de menu superior - selecione o dataset - botão "Visualização");
2. Adicione ao seu perfil os datasets de interesse para seu estudo: no formulário, escreva a justificativa e defina as palavras-chaves que descrevem a sua pesquisa (menu "Adicionar Datasets ao Perfil", na barra de menu superior - "Adicionar Datasets"- preencher formulário - botão "Confirmar").
3. Visualize os datasets adicionados ao seu perfil (menu "Meus Datasets", na barra de menu superior). Você visualizará a lista de datasets e a data de expiração do acesso.
4. A partir deste ponto você poderá baixar e utilizar o App PDD Desktop, a

A Plataforma de Dados Desidentificados (PDD) visa disponibilizar um ecossistema de dados para a produção de conhecimento em saúde, em especial sobre a pandemia de Covid-19, usando dados administrativos desidentificados individualizados e agregados. Para tal, identifica, obtém, organiza, processa e disponibiliza dados e informações que sejam úteis para realizar as análises pelos pesquisadores.

#### 1 Que tipo de dados você vai encontrar aqui?

Você poderá acessar dados administrativos de domínio público, abertos, não-nominais, agregados e microdados. São dados sobre informações demográficas e socioeconômicas, nutricionais, hospitalizações, nascimentos, óbitos, assistência e serviços de saúde, dados geoespaciais, além de casos, óbitos, testes diagnósticos e exames laboratoriais para Covid-19, vacinação contra Covid-19, medidas de prevenção e controle como o distanciamento social entre outros.

#### 2 Catálogo de metadados

Para cada conjunto de dados disponível na plataforma, existe um conjunto de características como fonte de informação, dicionário de variáveis, data de publicação, entre outros, ou seja, cada conjunto de dados possui seus próprios metadados, para melhor compreensão dos conteúdos.

Os metadados estão disponíveis no [Dataverse](#), que é um repositório de dados de pesquisas gratuito e de código-aberto, de fácil navegabilidade, que permite melhor gestão dos dados e metadados.

#### 3 Formato dos dados

Os dados podem ser baixados em formato CSV (Comma-separated values)

#### 4 Visualização de dados

No App PDD Desktop, após download do conjunto de dados, é possível explorá-los através de dashboard com algumas visualizações desses dados. Vale ressaltar que alguns conjuntos de dados disponíveis na PDD Cidacs possuem dashboards externos ao app PDD Desktop, e, nesses casos, como PNAD Contínua, IBP e IDS-Covid-19, um link estará disponível para direcioná-los para esses dashboards.

#### 5 Como ter acesso a estes dados?

É simples. Para ter acesso aos dados, primeiro adicione os datasets de interesse para seu estudo ao seu perfil. Então, baixe e instale o App PDD Desktop no seu computador para explorar e fazer download dos datasets selecionados. O passo a passo está logo abaixo.

##### Passo-a-passo para obter os dados:

1. Uma vez cadastrado e logado no portal Web da PDD, você poderá:

1.1) Acessar ao catálogo dos datasets disponíveis na plataforma. Ao clicar no botão "Datasets" na barra de menu superior, você obterá o conjunto de dados disponíveis. Ao clicar em cada dataset, você terá a descrição da base e uma amostra da estrutura dos dados;

1.2) Acessar o catálogo de metadados no Dataverse para obter a descrição, os dicionários e os identificadores persistentes das bases de dados (menu "Datasets", barra de menu superior - selecione o dataset - botão "Metadados");

1.3) Explorar os dados através da visualização de dados (menu "Datasets", barra de menu superior - selecione o dataset - botão "Visualização");

2. Adicione ao seu perfil os datasets de interesse para seu estudo: no formulário, escreva a justificativa e defina as palavras-chaves que descrevem a sua pesquisa (menu "Adicionar Datasets ao Perfil", na barra de menu superior - "Adicionar Datasets" - preencher formulário - botão "Confirmar").

3. Visualize os datasets adicionados ao seu perfil (menu "Meus Datasets", na barra de menu superior). Você visualizará a lista de datasets e a data de expiração do acesso.

4. A partir deste ponto você poderá baixar e utilizar o App PDD Desktop, a ferramenta de extração de dados, para obter os dados no seu computador. [Clique aqui](#) para saber mais.

A tela inicial é composta de blocos onde explicamos alguns detalhes do funcionamento e uso da aplicação web, eles são:

#### 1) Que tipo de dados você vai encontrar aqui?

Nesta seção discutimos quais tipos de dados estarão disponíveis na nossa plataforma.

#### 2) Catálogo de metadados

Nesta seção falamos sobre os metadados e comentamos brevemente sobre o nosso **Dataverse**, que será mais discutido ainda neste manual.

#### 3) Formato dos dados

Nesta seção falamos do formato dos dados disponibilizados.

#### 4) Visualização de dados

Nesta seção explicamos um pouco um formato de visualização destes dados a através da aplicação desktop.

#### 5) Como ter acesso a estes dados?

Nesta seção discutimos a forma de acesso aos dados, e disponibilizamos um passo a passo para auxiliar o usuário desde a exploração dos dados disponíveis, a requisição de uma base de dados e então o download desta base de dados através do nosso aplicativo desktop.

## IV. A barra de navegação

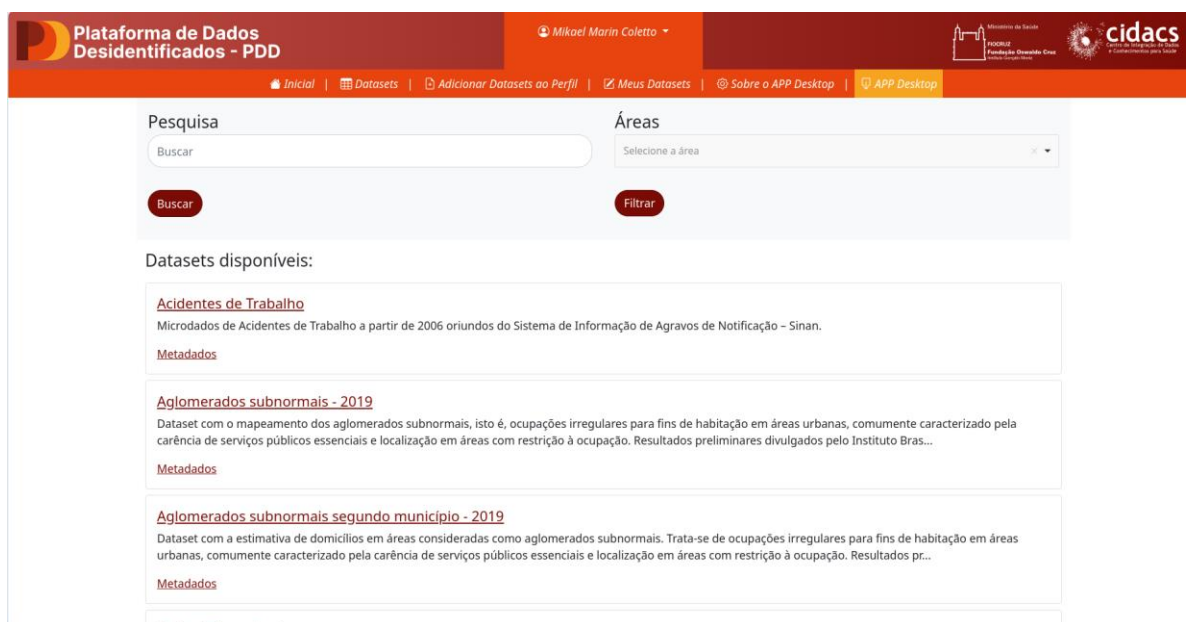


A barra de navegação é usada para alternar entre as áreas já comentadas da nossa aplicação web, apenas o último botão que foge deste comportamento e é um link direto para o download da nossa aplicação desktop.



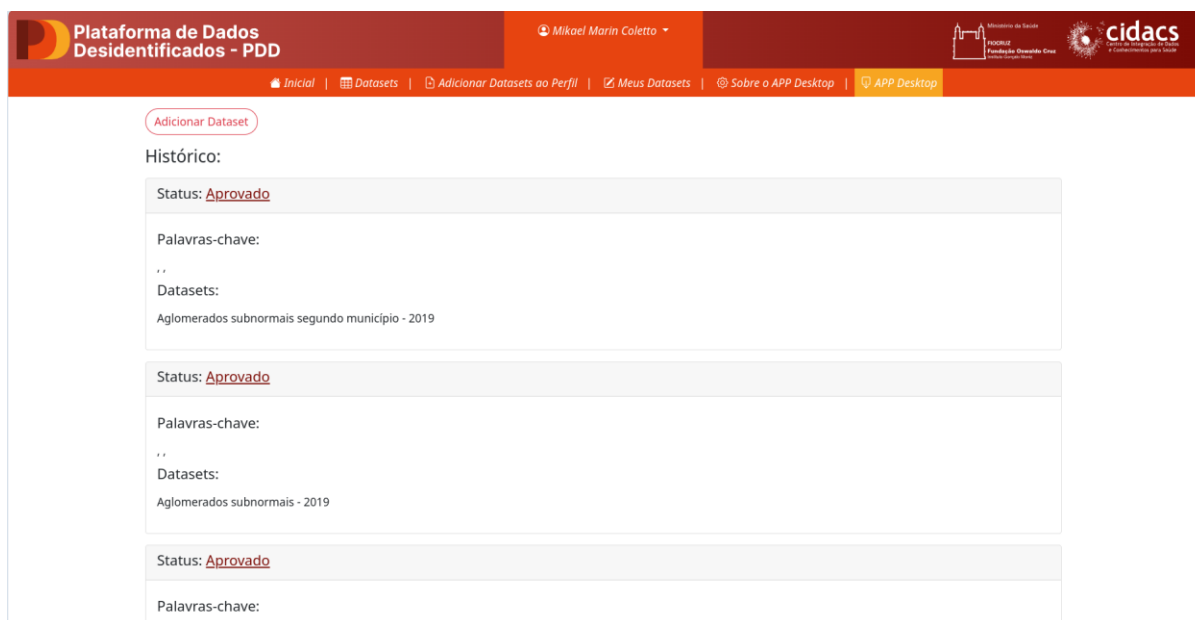
Além disso, no topo, temos o perfil do usuário, onde ao clicar podemos fazer algumas alterações no perfil ou sair da aplicação no nosso usuário.

## v. Datasets



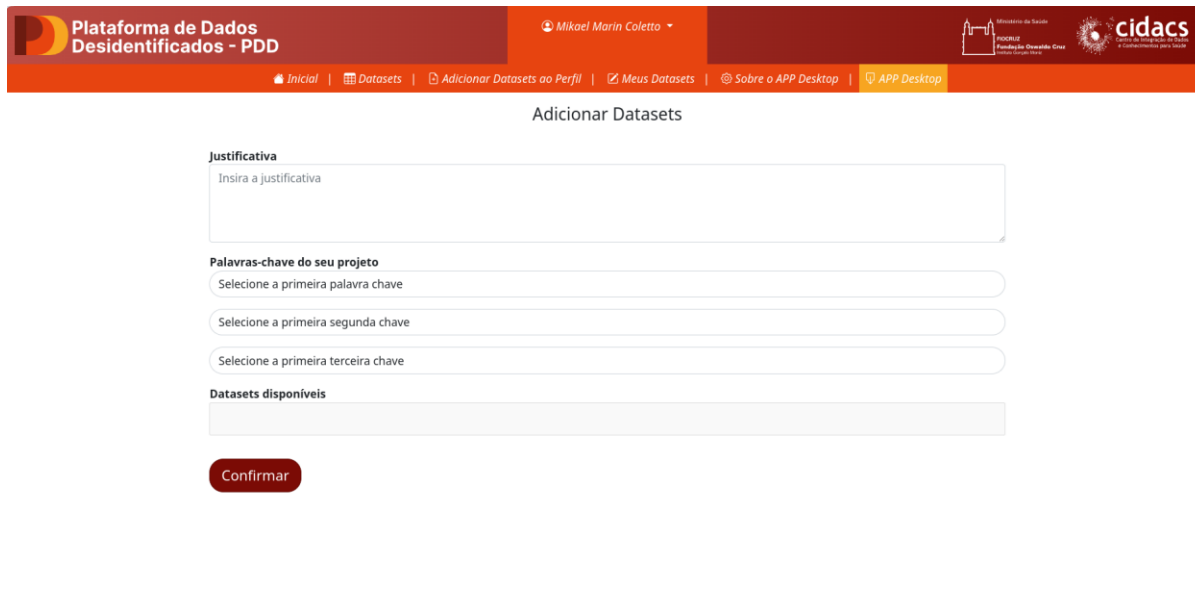
A tela dos datasets é composta de uma ferramenta de busca no topo, onde o usuário pode buscar um dataset diretamente digitando o nome ou parte do conteúdo deste dataset, pode buscar por área de interesse. E logo abaixo são listados os datasets disponíveis para aquela busca feita. Por padrão, antes de qualquer busca, todos os datasets disponíveis são listados.

## vi. Adicionar Datasets ao Perfil



The screenshot shows the top navigation bar of the 'Plataforma de Dados Desidentificados - PDD' interface. The user is logged in as 'Mikael Marin Coletto'. The navigation menu includes 'Inicial', 'Datasets', 'Adicionar Datasets ao Perfil', 'Meus Datasets', 'Sobre o APP Desktop', and 'APP Desktop'. A red button labeled 'Adicionar Dataset' is visible in the main content area.

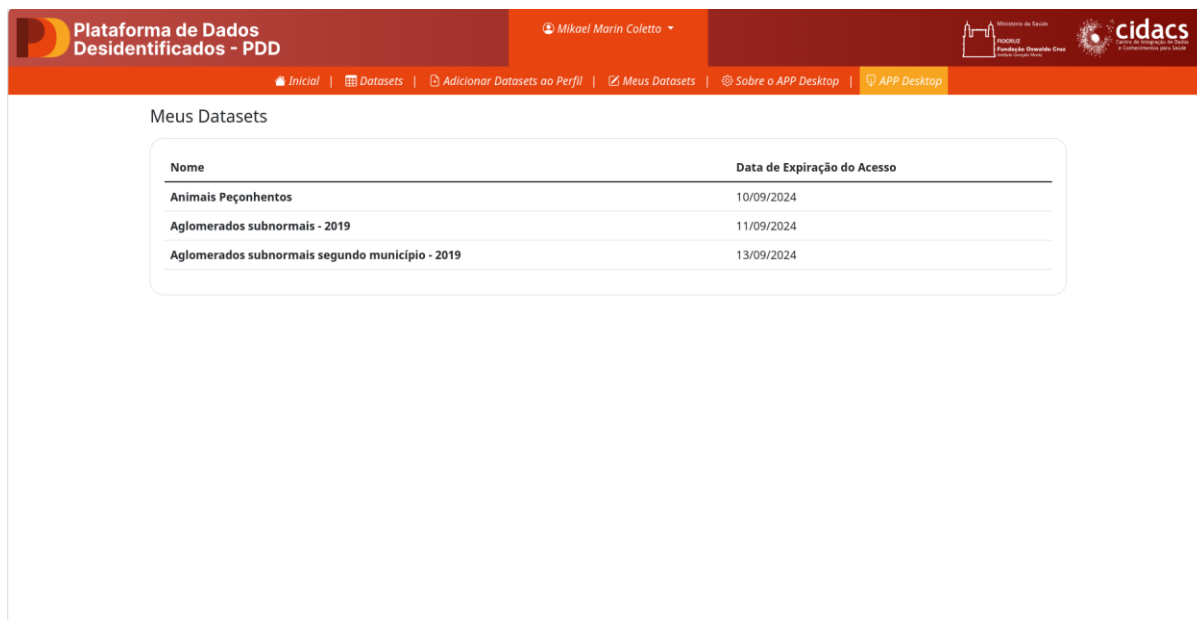
Nesta tela podemos ver um histórico das últimas requisições de bases de dados. No topo também vemos um botão onde é possível fazer a requisição de novos datasets para a conta. Ao clicar no botão de "adicionar Dataset" somos encaminhados para uma segunda tela.



The screenshot shows the 'Adicionar Datasets' form. It includes a 'Justificativa' field with a placeholder 'Insira a justificativa'. Below this are three 'Palavras-chave do seu projeto' fields with placeholders 'Selecione a primeira palavra chave', 'Selecione a primeira segunda chave', and 'Selecione a primeira terceira chave'. There is also a 'Datasets disponíveis' field. A red 'Confirmar' button is at the bottom.

Nesta tela temos os campos que precisam ser preenchidos, no mínimo um pequeno parágrafo com a justificativa e pelo menos uma palavra-chave, e então, no último campo, "Datasets disponíveis": podemos fazer a seleção das bases de dados que queremos solicitar.

## vii. Meus Datasets



Plataforma de Dados Desidentificados - PDD

Mikael Marin Coletto

Ministério da Saúde FIOCRUZ Fundação Oswaldo Cruz Instituto Gonçalo Moniz

cidacs Centro de Integração de Dados e Conhecimentos para Saúde

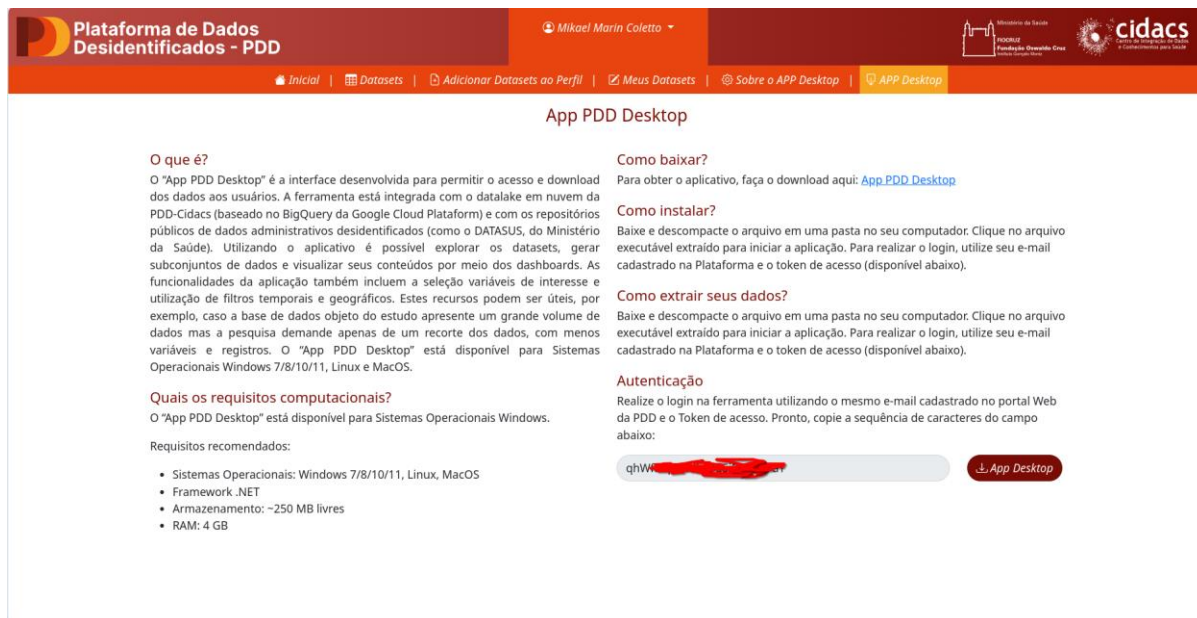
Inicial | Datasets | Adicionar Datasets ao Perfil | Meus Datasets | Sobre o APP Desktop | APP Desktop

### Meus Datasets

Nome	Data de Expiração do Acesso
Animais Peçonhentos	10/09/2024
Agglomerados subnormais - 2019	11/09/2024
Agglomerados subnormais segundo município - 2019	13/09/2024

Nesta tela podemos verificar quais datasets foram requeridos e qual o período de expiração do nosso pedido.

## viii. Sobre o APP Desktop



Plataforma de Dados Desidentificados - PDD

Mikael Marin Coletto

Ministério da Saúde FIOCRUZ Fundação Oswaldo Cruz Instituto Gonçalo Moniz

cidacs Centro de Integração de Dados e Conhecimentos para Saúde

Inicial | Datasets | Adicionar Datasets ao Perfil | Meus Datasets | Sobre o APP Desktop | APP Desktop

### App PDD Desktop

**O que é?**  
O "App PDD Desktop" é a interface desenvolvida para permitir o acesso e download dos dados aos usuários. A ferramenta está integrada com o datalake em nuvem da PDD-Cidacs (baseado no BigQuery da Google Cloud Platform) e com os repositórios públicos de dados administrativos desidentificados (como o DATASUS, do Ministério da Saúde). Utilizando o aplicativo é possível explorar os datasets, gerar subconjuntos de dados e visualizar seus conteúdos por meio dos dashboards. As funcionalidades da aplicação também incluem a seleção variáveis de interesse e utilização de filtros temporais e geográficos. Estes recursos podem ser úteis, por exemplo, caso a base de dados objeto do estudo apresente um grande volume de dados mas a pesquisa demande apenas de um recorte dos dados, com menos variáveis e registros. O "App PDD Desktop" está disponível para Sistemas Operacionais Windows 7/8/10/11, Linux e MacOS.

**Quais os requisitos computacionais?**  
O "App PDD Desktop" está disponível para Sistemas Operacionais Windows.

Requisitos recomendados:

- Sistemas Operacionais: Windows 7/8/10/11, Linux, MacOS
- Framework .NET
- Armazenamento: ~250 MB livres
- RAM: 4 GB

**Como baixar?**  
Para obter o aplicativo, faça o download aqui: [App PDD Desktop](#)

**Como instalar?**  
Baixe e descompacte o arquivo em uma pasta no seu computador. Clique no arquivo executável extraído para iniciar a aplicação. Para realizar o login, utilize seu e-mail cadastrado na Plataforma e o token de acesso (disponível abaixo).

**Como extrair seus dados?**  
Baixe e descompacte o arquivo em uma pasta no seu computador. Clique no arquivo executável extraído para iniciar a aplicação. Para realizar o login, utilize seu e-mail cadastrado na Plataforma e o token de acesso (disponível abaixo).

**Autenticação**  
Realize o login na ferramenta utilizando o mesmo e-mail cadastrado no portal Web da PDD e o Token de acesso. Pronto, copie a sequência de caracteres do campo abaixo:

qhW...

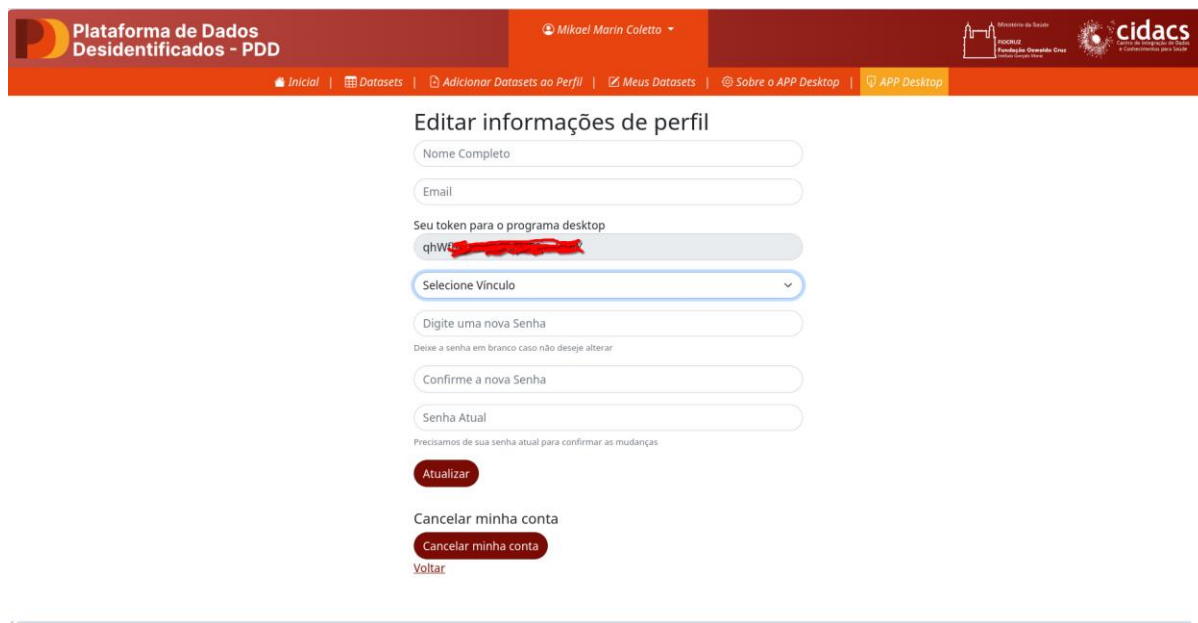
[App Desktop](#)

Aqui temos uma breve explicação da aplicação desktop e seus requisitos além de um resumo de como utilizá-la. Além disso, ao final, temos um token gerado que é único por usuário, utilizado para linkar a sua conta no nosso portal e na aplicação



web. Além disso, também temos uma segunda opção para download no botão ao lado do código de autenticação.

## IX. Editar Perfil



**Plataforma de Dados Desidentificados - PDD**

**Editar informações de perfil**

Nome Completo

Email

Seu token para o programa desktop

qhWt...

Selecione Vínculo

Digite uma nova Senha

Deixe a senha em branco caso não deseje alterar

Confirme a nova Senha

Senha Atual

Precisamos de sua senha atual para confirmar as mudanças

Atualizar

Cancelar minha conta

Cancelar minha conta

Voltar

Na tela de edição das informações de perfil podemos alterar o vínculo e/ou fazer uma alteração de senha, para isso vamos selecionar o novo vínculo ou digitar a nova senha, e precisamos da senha anterior para que seja feita a verificação de usuário e alteração dos dados.

Além disso, ao final, é possível cancelar a conta cadastrada clicando no botão de cancelar a conta. Lembrando que não é possível desfazer esta ação, mas é possível fazer um novo cadastro caso necessário.




### 3. Aplicação Desktop


A aplicação desktop é a ferramenta desenvolvida para consumo dos dados, possui interface gráfica e está disponível para os sistemas operacionais Windows, Linux e MacOS. Para sua utilização, é necessário baixá-la através do site, descompactar a pasta e então executá-la no computador do usuário.

Ela foi pensada para fornecer o acesso aos dados de forma mais simples e amigável e pode ser utilizada por qualquer tipo de usuário.

O arquivo baixado deve ser o mesmo abaixo:

Nome	Data de modificação	tipo	tamanho
▼ Hoje			
 PDD App Desktop	22/10/2024 09:04	Pasta compactada	394.704 KB
▼ Anteriormente neste ano			

E então o arquivo executável deve se parecer com este:

▼ Hoje			
 pdi-cidacs	22/10/2024 09:05	Aplicativo	168.256 KB

Após executar o arquivo, será pedido o nosso e-mail e o token para acessar a plataforma.



O e-mail é o mesmo que usamos no cadastro da aplicação web, no nosso site. Já o Token é o nosso identificador que é gerado ao efetuar o login e pode ser encontrado na aplicação web na aba "Sobre o APP Desktop". Referenciar o desktop (em caso de dúvidas)

Ao efetuarmos o login corretamente, a aplicação abrirá como a seguir:

1

**SELECIONE SEU DATASET**  

Sistema de Informação de Mortalidade (SIM) - a partir de 2020  
**Sistema de Informação Sobre Nascidos Vivos (SINASC) - a partir de 2020**  
Hepatites Virais  
Acidentes de Trabalho

2

**APLIQUE OS FILTROS (OPCIONAL)**  

UF Residência  
☒ Marcar todas  
PA, PB, PE, PI, PR, RJ, RN, RO, RR, RS, SC, SE, SP, TO  
Mês de Nascimento (AnoMês)

☐ CODMUNRES  
☒ ESCMAE  
☒ ESTCIVMAE  
☐ CODUFNATU  
☐ QTDGESTANT

3

SALVAR EM: C:/Users/Mikael/cidacs  
BAIXAR DADOS  
VISUALIZAR PRÉVIA DOS DADOS  
0%

**PRÉVIA DOS DADOS SELECIONADOS**

	APGARS	QTDFILVIVO	CONSULTAS	PARTO	PESO	CODMUNNASC	CONSPRENAT	ESCA
1			4	1	3000	150600	09	3
2	10	00	4	1	2850	410310	11	5
3	10	00	3	1	3392	350635	06	4
4		00	3	1	2800	510270	06	3
5	10	00	4	1	3405	351880	12	5
6	10	00	3	1	3470	352690	06	5
7	10		4	1	2986	280290	09	3
8	99	00	4	1	3760	355030	11	4
9	10	00	4	1	3350	330455	13	5
10	09	00	4	1	3084	411460	08	4
11		00	3	1	2770	150680	05	4
12		00	3	1	3500	130406	06	4
13		00	4	1	2600	293360	09	4
14	99	00	4	1	3000	355030	08	4

4

No canto superior esquerdo, temos uma caixa onde são apresentadas as bases de dados que solicitamos para download. No canto inferior esquerdo temos uma aba de filtros, abaixo dele temos um botão de download dos dados e um botão para visualizarmos uma prévia dos dados conforme os filtros selecionados, e por fim, no canto direito temos uma caixa onde será apresentada uma prévia dos dados conforme selecionado pelo usuário, e uma barra indicando o status de progresso da nossa solicitação. Abaixo explicaremos em detalhes cada um dos itens da aplicação.

1. Nesta caixa teremos todos os datasets que foram solicitados, podemos navegar entre eles e fazer o download individualmente, da base completa ou utilizando algum filtro apresentado abaixo.
2. A aplicação dos filtros é opcional e os filtros dependem da base selecionada, em geral podemos ter filtros relacionados ao período (ano ou ano mês), à região geográfica (em geral UFs), a idade ou outros fatores importantes referentes à base selecionada. Já à direita dos filtros podemos selecionar as variáveis que estão na base, podendo fazer o filtro de apenas algumas variáveis de interesse ou manter selecionadas todas para o download.
3. Abaixo temos um botão onde é indicado o caminho para onde serão transferidos os dados baixados, este caminho pode ser alterado ao clicar no próprio botão, ao lado temos os botões de download e por fim o botão de visualizar prévias. Ao clicarmos em download os dados serão baixados para o caminho apresentado logo acima, e ao clicar em prévias, após selecionarmos uma base e opcionalmente um filtro, teremos uma visualização resumida dos dados que estamos filtrando.
4. Neste campo de prévias podemos ver a base selecionada com os seus filtros aplicados, para observarmos melhor se é exatamente isso que estamos esperando antes de fazer o download dos dados.


Ao finalizarmos o processo teremos uma nova base de dados nos nossos arquivos, no caminho indicado pela aplicação, como no exemplo abaixo.

**Plataforma de Dados Desidentificados - PDD**

Ministério da Saúde  
FIOCRUZ  
Fundação Oswaldo Cruz  
Instituto Gonçalo Moniz

**cidacs**  
Centro de Integração de Dados e Conhecimentos para Saúde

 > Este Computador > Disco Local (C:) > Usuários > Mikael > cidacs

	Nome	Data de modificação	Tipo	Tamanho
al	 sinasc.csv	22/10/2024 10:51	Arquivo CSV	15.917 KB
-				

## 4. API da PDD

Além da aplicação Desktop, nós também desenvolvemos outra forma de acesso aos dados utilizando uma API em python, o objetivo é que torne o processo mais rápido e possível de ser automatizado para aqueles que preferem o uso através do uso de códigos de programação. Porém, ela requer um conhecimento um pouco maior do funcionamento de linguagens de programação de uso de APIs.

O acesso via API possui algumas vantagens, além da possível automatização da rotina, temos a utilização direta do nosso banco de dados calcular as consultas específicas, diminuindo o consumo de memória e processamento do usuário, uso de linguagem SQL para criação de consultas particulares e mais algumas utilidades próprias da biblioteca para analisar as bases de dados.

Para utilizarmos a biblioteca, é preciso primeiramente fazer o download do arquivo zip no seu computador. Este arquivo conterá a pasta da biblioteca, um script no formato de Jupiter Notebook e um arquivo adicional contendo uma apresentação da biblioteca com suas funcionalidades e recursos.

Downloads > PDD API				
	Nome	Data de modificação	Tipo	Tamanho
	Hoje			
	pdd-cidacs-lib-main	06/11/2024 10:52	Pasta de arquivos	
	Último mês			
	Tutorial PDDCidacs Lib.ipynb	24/10/2024 15:29	Arquivo IPYNB	42 KB
	PDD API Cidacs Apresentação	23/10/2024 09:59	Adobe Acrobat D...	1.677 KB

A primeira pasta contém os arquivos da biblioteca que serão utilizados.

O arquivo de tutorial é no formato `.pynb` e pode ser utilizado através de IDEs que leiam este formato de notebook e tenham acesso ao kernel python da máquina.

Este arquivo contém um passo à passo de como utilizar a ferramenta e pode ser utilizado diretamente ou sendo alterado nos pontos que o usuário necessitar. Ele apresenta um uso de caso contendo as principais funcionalidades da aplicação e algumas formas de uso da API. Ressaltamos que para o acesso aos dados, ainda é preciso um cadastro na nossa aplicação web e a utilização do seu token pessoal.

Listaremos algumas destas funcionalidades disponíveis no lançamento da API. As demais podem ser vistas no script ou utilizando o método "help" da biblioteca.

Para o seu uso, é necessário primeiro importar a biblioteca e então utilizar o método de ajuda como abaixo:

```
#importando a biblioteca
from bib_pdd_cidacs import BibPddCidacs

#instanciando um cliente
client = BibPddCidacs()

help(client)
```

Temos uma função que é utilizada para autenticar e permitir a conexão entre o usuário e o nosso banco de dados da PDD (**authentication**).

```
credentials = {
    "email": "meuemail@email.com",
    "token": "meutoken"
}

client.authentication(credentials)
```

Depois podemos usar a função para listar os bancos (**list\_db**) de dados que estão atrelados à nossa conta (para mais informações sobre como adicionar bancos de dados à conta, consultar [Adicionar Datasets ao Perfil](#)).

```
dbs = client.list_db()

print('Meus Datasets:')
for cod in sorted(list(dbs.keys())):
    print(f'Código: {cod}', f'- Dataset: {dbs[cod]}')
```

Podemos observar o formato de um dataset que está no nosso perfil, utilizando a função **shape**. Para este exemplo usaremos o dataset de código: 4 - Sistema de Informação Sobre Nascidos Vivos (SINASC) - a partir de 2020.

```
# Sistema de Informação Sobre Nascidos Vivos (SINASC) - a partir de 2020
dataset = 4
quant_registros = client.shape(dataset=dataset)
print(f'Qt. de registros: {quant_registros[0]}, Qt. de variáveis {quant_registros[1]}')
```

Também podemos verificar quais colunas estão disponíveis neste dataset através do comando **list.columns**.

```
dataset = 4 # SINASC
lista_nome_variaveis = client.list_columns(dataset=dataset)
lista_nome_variaveis
```

É possível ainda realizar diretamente uma consulta (*query*) em linguagem SQL com o comando **query**.

```
dataset = 240
query = f"SELECT __PDD_DTOBITO_ANO as ANO, CODMUNRES, SEXO, COUNT (*) AS OBITO \
FROM {dataset} WHERE REGEXP_CONTAINS(CAUSABAS, r'^I(2[0-5]||[3-5][0-9])') \
AND __PDD_DTOBITO_ANO >= '2000' AND __PDD_DTOBITO_ANO <= '2019' \
AND IDADE >= '25' AND IDADE <= '35' \
GROUP BY __PDD_DTOBITO_ANO, CODMUNRES, SEXO"

r_fx_25a35 = client.query(query = query)

Query = f"select
Client.download(query='select ANO, CODMUNRES, SEXO from 240 WHERE UF = 'Bahia'',
filename='sinasc.csv')
```

O script de modelo possui mais alguns exemplos e pode ser consultado para maiores esclarecimentos. Além dos exemplos, o script terá comentários explicando suas funções e o como fazer utilização delas, servindo também como um manual mais detalhado e prático do uso da nossa API.

## 5. Metadados (Dataverse)

### I. A plataforma de Governança de Dados

No ambiente contemporâneo, onde a geração e o compartilhamento de dados científicos e acadêmicos estão em constante expansão, a governança de dados tornou-se essencial para garantir a integridade, a acessibilidade e a reutilização eficaz das informações. Nesse contexto, a plataforma **Dataverse Project** surge como uma solução robusta e colaborativa, desenvolvida para facilitar o armazenamento, a gestão e a disseminação de dados de pesquisa, promovendo uma governança de dados eficiente e alinhada às melhores práticas internacionais.

#### a. Dataverse

O Dataverse é uma plataforma de código aberto destinada ao compartilhamento, citação, análise e preservação de dados de pesquisa. Inicialmente desenvolvido pela Harvard University, o projeto agora conta com uma comunidade global de instituições acadêmicas, organizações de pesquisa e desenvolvedores que contribuem para sua evolução contínua. Através de uma interface amigável e de recursos avançados, o Dataverse permite que pesquisadores publiquem seus conjuntos de dados de maneira organizada e acessível, facilitando a colaboração e a transparência científica.

Os principais componentes utilizados são:

- **Dataverses (Conjuntos de Dados):** Estruturas organizacionais que permitem agrupar e categorizar conjuntos de dados relacionados, facilitando a navegação e a descoberta de informações dentro da plataforma.
- **Metadados Rigorosos:** Utilização de padrões de metadados para descrever detalhadamente os conjuntos de dados, assegurando que informações relevantes sobre a origem, a metodologia e o conteúdo estejam claramente documentadas.
- **Controle de Acesso e Permissões:** Mecanismos robustos que permitem aos proprietários dos dados definir níveis de acesso, garantindo que apenas usuários autorizados possam visualizar ou modificar os conjuntos de dados conforme as políticas estabelecidas.
- **Preservação e Backup:** Infraestrutura que garante a preservação a longo prazo dos dados, com backups regulares e medidas de segurança para proteger contra perda ou corrupção de informações.

#### b. A nossa Governança de Dados

A governança de dados refere-se ao conjunto de práticas, políticas e processos que asseguram a gestão eficiente e responsável dos dados dentro de uma



organização ou comunidade. Nós implementamos, através da ferramenta Dataverse essa governança através dos seguintes aspectos:

- **Padronização de Metadados:** Ao adotar padrões reconhecidos internacionalmente para a descrição de dados, a plataforma assegura a consistência e a interoperabilidade dos conjuntos de dados, facilitando a integração e o compartilhamento entre diferentes disciplinas e instituições.
- **Transparência e Reprodutibilidade:** A disponibilização pública dos dados, acompanhada de documentação detalhada, promove a transparência nas pesquisas e permite que outros pesquisadores reproduzam e validem os resultados, fortalecendo a integridade científica.
- **Segurança e Conformidade:** Com recursos avançados de controle de acesso e proteção de dados, o Dataverse Project ajuda as organizações a cumprir com regulamentações de privacidade e proteção de dados, como a GDPR (Regulamento Geral sobre a Proteção de Dados) e outras legislações pertinentes.
- **Gestão de Ciclo de Vida dos Dados:** A plataforma facilita o gerenciamento do ciclo de vida dos dados, desde a sua criação e coleta até o arquivamento e a preservação a longo prazo, garantindo que os dados permaneçam acessíveis e utilizáveis ao longo do tempo.
- **Responsabilidade e Auditoria:** Ferramentas de monitoramento e registro de atividades permitem rastrear alterações e acessos aos conjuntos de dados, promovendo a responsabilidade e facilitando auditorias de conformidade e integridade dos dados.

## II. O Dataverse do CIDACS

Com o intuito de construir uma plataforma de dados desidentificados voltada para a disponibilização do acesso a pesquisadores e ao público em geral, o CIDACS criou seu próprio Dataverse para lidar com a Governança e Gestão dos dados, nele podemos encontrar diversos catálogos de diferentes projetos pertencentes ao CIDACS. Dentro da nossa plataforma da PDD, utilizamos a arquitetura criada e gerimos um catálogo de dados específico para o nosso projeto.

 **Dataverse**

Search ▾ User Guide Support Sign Up Log In

 **Ministério da Saúde**  
**FIOCRUZ**  
**Fundação Oswaldo Cruz**  
Instituto Gonçalo Moniz

 **cidacs**

**Dataverse**

**Centro de Integração de Dados e Conhecimento para Saúde**  
(CIDACS/Fiocruz)

 Metrics

562 Downloads

 Contact  Share


CIDACS/Fiocruz-BA data repository publishes epidemiological data and research from our researchers and projects


 **Plataforma de Dados Desidentificados - PDD**  
PDD Cidacs - Plataforma de Dados Desidentificados Cidacs


 **NIHR - Grupo de pesquisa em Saúde Global sobre Políticas e Desigualdade em Saúde**


 **Projeto Gates - A coorte de 100 milhões de brasileiros**

 **Plataforma de vigilância de longo prazo para a Zika e Microcefalia no âmbito do SUS**

Search this dataverse...  **Advanced Search**

☒  **Dataverses (6)**


☒  **Datasets (119)**

☐  **Files (265)**

**Dataverse Category**  
Research Project (6)

**Publication Year**

**1 to 10 of 125 Results**





**Replication Data for: Sistema de Informação de Agravos de Notificação - SINAN**  
Aug 12, 2024 - PDD Cidacs - Plataforma de Dados Desidentificados Cidacs  
Ministério da Saúde; Secretaria de Vigilância em Saúde, 2024, "Replication Data for: Sistema de Informação de Agravos de Notificação - SINAN", <https://doi.org/10.57833/cidacs/BWAF4P>, Centro de Integração de Dados e Conhecimento para Saúde, V2  
Este conjunto de dados possui as doenças de notificação compulsória que são registradas no Sistema de Informação de Agravos de Notificação - Sinan. O Sinan é alimentado, principalmente, pela notificação e investigação de casos de doenças e agravos que constam da lista nacional de...

A nossa tela inicial já direciona o usuário a observar todos os bancos de dados cadastrados e que estão à disposição na nossa aplicação da PDD. Ainda é possível efetuar buscas diretamente no filtro, fazer buscas avançadas ou utilizar alguns filtros, como ano de publicação ou área do conhecimento.

 **Dataverse**

Search ▾ User Guide Support Sign Up Log In

 **Ministério da Saúde**  
**FIOCRUZ**  
**Fundação Oswaldo Cruz**  
Instituto Gonçalo Moniz


 **cidacs**


**Dataverse**


**Centro de Integração de Dados e Conhecimento para Saúde**  
(CIDACS/Fiocruz)


 Metrics


562 Downloads

 Contact  Share

Search this dataverse...  **Advanced Search**

☒  **Dataverses (6)**


☒  **Datasets (119)**

☐  **Files (265)**


**Dataverse Category**  
Research Project (6)

**Publication Year**  
2024 (6)  
2023 (2)  
2022 (118)

**Subject**  
Medicine, Health and Life Sciences (124)  
Social Sciences (112)  
Earth and Environmental Sciences (2)  
Other (2)  
Mathematical Sciences (1)

**Publication Status: Published** 


**1 to 10 of 125 Results**



**Replication Data for: Sistema de Informação de Agravos de Notificação - SINAN**  
Aug 12, 2024 - PDD Cidacs - Plataforma de Dados Desidentificados Cidacs  
Ministério da Saúde; Secretaria de Vigilância em Saúde, 2024, "Replication Data for: Sistema de Informação de Agravos de Notificação - SINAN", <https://doi.org/10.57833/cidacs/BWAF4P>, Centro de Integração de Dados e Conhecimento para Saúde, V2  
Este conjunto de dados possui as doenças de notificação compulsória que são registradas no Sistema de Informação de Agravos de Notificação - Sinan. O Sinan é alimentado, principalmente, pela notificação e investigação de casos de doenças e agravos que constam da lista nacional de...



**Replication For: Relatório de pagamentos do bolsa família por município e ano/mês**  
Aug 12, 2024 - PDD Cidacs - Plataforma de Dados Desidentificados Cidacs  
Ministério do Desenvolvimento e Assistência Social, Família e Combate à Fome (MDS), 2024, "Replication For: Relatório de pagamentos do bolsa família por município e ano/mês", <https://doi.org/10.57833/cidacs/NPH-MJO>, Centro de Integração de Dados e Conhecimento para Saúde, V1, UNF:6.1noxm1SDG17JeyLP7LWJOA= [fileUNF]  
Este dataset consiste no número de famílias beneficiárias do programa bolsa família, valor total e médio repassado agregado por município e ano/mês a partir de 2004.



**Replication Data for: Programa Auxílio Brasil**  
Jul 25, 2024 - PDD Cidacs - Plataforma de Dados Desidentificados Cidacs  
Ministério do Desenvolvimento e Assistência Social, Família e Combate à Fome, 2024, "Replication Data for: Programa Auxílio Brasil", <https://doi.org/10.57833/cidacs/N1MAPC>, Centro de Integração de Dados e Conhecimento para Saúde, V1, UNF:6.YASvnmPqCpEpEMp082zg= [fileUNF]


## a. Busca


Logo na primeira página podemos observar os campos de busca e filtros que podem ser utilizados para encontrar um dataset ou metadado.


Podemos buscar por ano de adição dos metadados ao nosso Dataverse, por área do conhecimento e ainda é possível efetuar uma busca avançada.

[Advanced Search](#)

---

☒  **Dataverses (0)**


☒  **Datasets (95)**

☐  **Files (249)**

**Publication Year**  
2024 (5)  
2023 (2)  
2022 (88)


**Subject**  
Medicine, Health and Life Sciences (95)  
Social Sciences (94)  
Earth and Environmental Sciences (1)  
Other (1)

**1 to 10 of 95 Results**



**Replication Data for: Sistema de**  
Aug 12, 2024  
Ministério da Saúde; Se  
- SINAN", <https://doi.org>

Este conjunto de dados possui as d  
Sinan. O Sinan é alimentado, princi



**Replication For: Relatório de paç**  
Aug 12, 2024  
Ministério do Desenvolv  
pagamentos do bolsa fa  
Conhecimento para Saú

Este dataset consiste no número de  
ano/mês a partir de 2004.

A busca avançada pode ser feita por informações presentes no nome, identificador, afiliação, descrição, área do conhecimento ou outros metadados.

The screenshot shows the Dataverse search interface. At the top, there's a navigation bar with 'Dataverse' logo and links for Search, User Guide, Support, Sign Up, and Log In. Below this, a breadcrumb trail reads 'Centro de Integração de Dados e Conhecimentos para Saúde > PDD Cidacs - Plataforma de Dados Desidentificados Cidacs >'. A 'Find' button is on the left. The main search area is divided into two sections: 'Dataverses' and 'Datasets: Citation Metadata'. Both sections have input fields for Name, Identifier, Affiliation, and Description. The 'Subject' field is a dropdown menu with options like Agricultural Sciences, Arts and Humanities, Astronomy and Astrophysics, Business and Management, and Chemistry.

Ao fazer uma busca de exemplo, no nosso caso, IBP, buscando sobre o Índice Brasileiro de Privação desenvolvido pelo CIDACs, encontramos duas opções que podem ser selecionadas.

The screenshot shows the search results for 'ibp'. On the left, there's a sidebar with filters for 'Dataverses (0)', 'Datasets (2)', and 'Files (0)'. Below these are filters for 'Publication Year' (2022 (2)) and 'Subject' (Medicine, Health and Life Sciences (2), Social Sciences (2)). The main results area shows '1 to 2 of 2 Results'. The first result is 'Índice Brasileiro de Privação (IBP) Municipal' dated Jul 7, 2023, with a description mentioning the Centro de Integração de Dados e Conhecimentos para Saúde and Instituto Brasileiro de Geografia e Estatística. The second result is 'Índice Brasileiro de Privação (IBP) por Setor Censitário' dated Dec 15, 2022, with a similar description. Both results include an 'Alternative URL'.



## b. Informações de um dataset

Ao selecionarmos uma das bases, no nosso exemplo, o Índice Brasileiro de Privação (IBP) Municipal, somos direcionados para um resumo do dataset, seu doi e forma de citação e podemos acessar os arquivos referentes a sua documentação e dicionário

dataverse.cidacs.org/dataset.html

Search User Guide Support Sign Up Log In

PDD Cidacs - Plataforma de Dados Desidentificados Cidacs  
(cidacs.bahia.fiocruz.br)

Centro de Integração de Dados e Conhecimento para Saúde > PDD Cidacs - Plataforma de Dados Desidentificados Cidacs >

## Índice Brasileiro de Privação (IBP) Municipal

Version 3.0

Centro de Integração de Dados e Conhecimentos para Saúde; Instituto Brasileiro de Geografia e Estatística, 2022, "Índice Brasileiro de Privação (IBP) Municipal", <https://doi.org/10.57833/cidacs/A5JAHB>, Centro de Integração de Dados e Conhecimento para Saúde, V3, UNF:6:CaLiDNZz8EDVY8FQE4e8w== [fileUNF]

Cite Dataset - Learn about Data Citation Standards.

Access Dataset -  
Contact Owner Share

Dataset Metrics  
65 Downloads

Description  
Dados do Índice Brasileiro de Privação (IBP) calculado para os municípios brasileiros. Representa uma medida de privação material ou, de um modo mais geral, níveis de posição socioeconômica. Este índice foi calculado utilizando informações de renda, educação e condições de moradia a partir dos dados do Censo Demográfico 2010. Desenvolvido por pesquisadores do Centro de Integração de Dados e Conhecimentos para Saúde (Cidacs/Fiocruz Bahia) e da Universidade de Glasgow-Escócia, dentro do projeto Social Policy & Health Inequalities (SPHI), financiado pela agência de fomento do Reino Unido National Institute for the Health Research (NIHR).

Subject  
Medicine, Health and Life Sciences; Social Sciences

License/Data Use Agreement  
CC BY 4.0

Files Metadata Terms Versions

Search this dataset...

Filter by  
File Type: All Access: All

1 to 3 of 3 Files Download

Inicialmente vemos o nome do dataset, acompanhado da sua forma de citação, doi, e uma pequena descrição.

Centro de Integração de Dados e Conhecimento para Saúde > PDD Cidacs - Plataforma de Dados Desidentificados Cidacs >

## Índice Brasileiro de Privação (IBP) Municipal

Version 3.0



Centro de Integração de Dados e Conhecimentos para Saúde; Instituto Brasileiro de Geografia e Estatística, 2022, "Índice Brasileiro de Privação (IBP) Municipal", <https://doi.org/10.57833/cidacs/A5JAHB>, Centro de Integração de Dados e Conhecimento para Saúde, V3, UNF:6:CaLiDNZz8EDVY8FQE4e8w== [fileUNF]

Cite Dataset -

Learn about Data Citation Standards.

Access Dataset -

Contact Owner

Share

Dataset Metrics

68 Downloads

### Description

Dados do Índice Brasileiro de Privação (IBP) calculado para os municípios brasileiros. Representa uma medida de privação material ou, de um modo mais geral, níveis de posição socioeconômica. Este índice foi calculado utilizando informações de renda, educação e condições de moradia a partir dos dados do Censo Demográfico 2010. Desenvolvido por pesquisadores do Centro de Integração de Dados e Conhecimentos para Saúde (Cidacs/Fiocruz Bahia) e da Universidade de Glasgow-Escócia, dentro do projeto Social Policy & Health Inequalities (SPHI), financiado pela agência de fomento do Reino Unido National Institute for the Health Research (NIHR).

### Subject

Medicine, Health and Life Sciences; Social Sciences

### License/Data Use Agreement



CC BY 4.0

No botão de citação existem alguns formatos compatíveis que podem ser exportados.

[Centro de Integração de Dados e Conhecimento para Saúde](#) > [PDD Cidacs - Plataforma de Dados Desidentificados Cidacs](#) >

## Índice Brasileiro de Privação (IBP) Municipal

Version 3.0



Centro de Integração de Dados e Conhecimentos para Saúde; Instituto Brasileiro de Geografia e Estatística, 2022, "Índice Brasileiro de Privação (IBP) Municipal", <https://doi.org/10.57833/cidacs/A5JAHB>, Centro de Integração de Dados e Conhecimento para Saúde, V3, UNF:6:CaLiDNZz8EDVY8FQE4e8w== [fileUNF]

Cite Dataset ▾

Learn about [Data Citation Standards](#).

EndNote XML

RIS

BibTeX

Description ?

O Índice Brasileiro de Privação (IBP) calculado para os municípios brasileiros. Representa uma medida de privação material ou, de um modo mais geral, níveis de posição socioeconômica.

Logo abaixo vemos os arquivos disponibilizados referentes a este dataset. Podemos baixá-los individualmente ou baixar todos em um único arquivo zip.

[Files](#) [Metadata](#) [Terms](#) [Versions](#)

Search this dataset...

Filter by  
File Type: All ▾ Access: All ▾

Sort ▾

1 to 3 of 3 Files

☐

DataDocumentation\_IBP.pdf  
Adobe PDF - 849.9 KB  
Published Dec 15, 2022  
26 Downloads

☐

IBP\_Municipalities Level\_Short.tab  
Tabular Data - 428.0 KB  
Published Jul 7, 2023  
18 Downloads  
9 Variables, 5565 Observations

☐

metadados\_IBP\_Municipal.pdf  
Adobe PDF - 73.9 KB  
Published Jan 28, 2023  
24 Downloads



Na segunda aba, podemos observar com mais detalhes os metadados cadastrados, como o doi do metadado, o autor dos dados, uma breve descrição e a data de publicação dos metadados na nossa plataforma..

The screenshot shows the Dataverse interface for a dataset. The top navigation bar includes 'Search', 'User Guide', 'Support', 'Sign Up', and 'Log In'. The dataset title is 'Índice Brasileiro de Privação (IBP) Municipal'. The description states it is a measure of material deprivation calculated for Brazilian municipalities. The subject is 'Medicine, Health and Life Sciences; Social Sciences'. The license is 'CC BY 4.0'. The citation metadata section includes the Dataset Persistent ID (doi:10.57833/cidacs/A5JAHB), Publication Date (2022-12-15), Title, Alternative URL, Author (Centro de Integração de Dados e Conhecimentos para Saúde), Contact, and Description. There is an 'Export Metadata' button and tabs for 'Files', 'Metadata', 'Terms', and 'Versions'.

Já na terceira e quarta aba, temos respectivamente as informações referentes aos termos de uso dos metadados e às atualizações e versões destes metadados.

The screenshot shows the 'Terms' and 'Versions' tabs of the dataset page. The title is 'Índice Brasileiro de Privação (IBP) Municipal'. The description is the same as in the previous screenshot. The subject is 'Medicine, Health and Life Sciences; Social Sciences'. The license is 'CC BY 4.0'. The 'Versions' tab shows a table with the following data:

Dataset Version	Summary	Contributors	Published on
3.0	Files (Added: 1); View Details	Rafael Souza	2023-07-07
2.0	Files (Added: 1); View Details	Rafael Souza	2023-01-28
1.0	This is the first published version.	Rafael Souza, Jonas Silva do Espírito Santo	2022-12-15

There is a 'View Differences' button and tabs for 'Files', 'Metadata', 'Terms', and 'Versions'.



### c. Citação dos dados

Para fazer uma citação, recomendamos que cite a plataforma de dados desidentificados como fonte onde os dados foram coletados.