

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D15W40>

Title: Distribuição dos Garbage Codes no sistema de Informação de Mortalidade, Brasil, 2000 a 2020

Creator: Silvia lucia Borowicc - **ORCID:** [0000-0001-7399-274X](https://orcid.org/0000-0001-7399-274X)

Affiliation: Universidade de São Paulo (www5.usp.br)

Principal Investigator: Érika Carvalho De Aquino

Project Administrator: Solange Nice Alves-Souza, Otaliba Libanio Morais Neto

Funder: Universidade de São Paulo (www5.usp.br)

Template: Digital Curation Centre (português)

Start date: 03-01-2022

End date: 03-24-2023

Last modified: 01-18-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Distribuição dos Garbage Codes no sistema de Informação de Mortalidade, Brasil, 2000 a 2020

Fontes de dados:

1. Registros de mortalidade no Brasil entre os anos de 2000 e 2020;
2. CID10 - Código Internacional de Doenças;
3. Regiões, estados e municípios do Brasil;

Fontes de dados:

1. SIM: <ftp.datasus.gov.br/dissemin/publicos/SIM/CID10/DORES>
2. CID10: ftp.datasus.gov.br/dissemin/publicos/SIHSUS/200801_/Auxiliar/TAB_SIH.zip
3. Município:
 1. www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?=&t=downloads
 2. geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_municipais/municipio_2020/Brasil
4. População: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?ibge/cnv/popsvsbr.def>

Coleta e tratamento:

Tratamento realizado utilizando a ferramenta Apache Hop, cujos arquivos estão disponíveis no repositório, pasta <codigo_fonte>.

1. SIM:
 1. Realizado tratamento de data de óbito nula ou inválida atribuindo valor padrão com o ano obtido do arquivo de dados do SIM, dia e mês representados com o número 0 já que não pudemos determinar data exata;
 2. Faixa-etária, sexo e raça/cor enriquecidas por mapeamento conforme o dicionário de dados do TABNET;
 1. dado inválido ou não informado: mapeado para "Ignorado" ou "Ignorada";
 3. Tratamento da idade a partir do código, de acordo com o dicionário de dados do TABNET;
 2. CID10:
 1. Dados enriquecidos com as descrições dos códigos e grupos da classificação CID10;
 2. Códigos CID10 associados à tabela de Garbage Codes Gbd 2017.
 3. Município:
 1. Dados enriquecidos com dados associados aos municípios, permitindo agregações por estado e região;
 2. Dados do município enriquecidos com coordenadas geográficas.
 4. População:
 1. Incorporados dados da população municipal para permitir cálculo de taxa.
1. Dicionário de dados (arquivos identificados pelo nome do arquivo de dados com o prefixo dic_);
 2. Modelo de dados da base integrada;
 3. Arquivos das implementações de coleta, transformação e consolidação dos dados;
 4. Informações sobre acesso ao repositório;

Não se aplica pois foram utilizados dados abertos.

Não se aplica pois foram utilizados dados abertos e também será aberto o conjunto de dados deste plano.

Os dados serão armazenados no repositório da USP.

O repositório é gerenciado pela USP.

Não se aplica.

Não se aplica.

Os dados serão depositado em repositório da USP.

O acesso é aberto para consulta.

O autor do plano de dados.

Disponibilidade do repositório de dados da USP.

Planned Research Outputs

Dataset - "Distribuição dos Garbage Codes"

[Acesse o Repositório da USP](#)

Pastas:

- garbage_dados: conjuntos de dados tratados, e seus respectivos dicionários;
 - proporcao: dados da proporção de óbitos identificados com garbage codes em relação ao total de óbitos;
 - taxa: taxa bruta e padronizada de óbitos identificados com garbage codes;
 - visualizacao: gráficos para visualização das informações baseadas nos conjuntos de dados;
- garbage_etl: arquivos da aplicação Apache Hop com a implementação do fluxo de tratamento e carga de dados;
 - dados: tabela CID10 e lista Garbage codes GBD2017;
 - modelo_relacional: modelo da base de dados resultante do tratamento realizado.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Distribuição dos Garbage Codes	Dataset	2023-03-30	Open	None specified		Creative Commons Attribution 4.0 International	None specified	No	No