

Plan Overview

A Data Management Plan created using DMPTool

Title: Construção de um banco de dados a partir da base do Censo da Educação Superior - Docentes

Creator: Roberta França - ORCID: [0000-0003-0992-9652](https://orcid.org/0000-0003-0992-9652)

Affiliation: Universidade Federal de São Paulo (unifesp.br)

Funder: Digital Curation Centre (dcc.ac.uk)

Template: Digital Curation Centre

Project abstract:

O Centro de Estudos e Think Tank Sociedade, Universidade e Ciência- SOU_Ciência, vinculado à Universidade Federal de São Paulo, tem como um de seus objetivos possibilitar aos seus pesquisadores, e/ ou aos pesquisadores interessados, o acesso a série histórica do banco de dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) sobre os docentes das Instituições de Ensino Superior (IES). Sendo que estes dados foram tratados, reorganizados e reapresentados permitindo correlacioná-los à temática da expansão do ensino superior no Brasil e/ou cruzar essas informações com dados de outras bases e fomentar outras pesquisas no campo da educação. Os dados trabalhados representam o período histórico de 2009 a 2019. A metodologia envolveu também a compatibilização dos códigos do banco do Censo da Educação Superior, uma vez que por motivos de reconfiguração de política pública educacional no período analisado o Inep adotou metodologias diferentes tanto para o processo de obtenção das informações, quanto para o conteúdo da informação disponibilizada. A metodologia e documentação necessária para sua reprodução segue apresentada, bem como o banco de dados. O intuito é possibilitar ao pesquisador compreender a informação, os denominadores comuns e entrecruzar os dados para comparação entre Índices ou qualquer outra intenção de pesquisa científica ou acesso a informação.

Start date: 12-03-2021

End date: 07-25-2022

Last modified: 05-10-2023

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

Construção de um banco de dados a partir da base do Censo da Educação Superior - Docentes

Mapeamento das IES - é documento em .xls, que traz as informações das Universidades a partir de dados selecionados do Dicionário de Variáveis e do banco de Microdados do Inep, no período de 2009 a 2019.

Tabela com os Dados Administrativos das IES traduzidos (2009 -2019), ou seja, os códigos do IBGE planejados com as informações traduzidas.

Fluxo dos Processos de Extração, Transformação e Carga - em .xls, elaborado utilizando a ferramenta Pentaho Data Integration (PDI), que é um software livre.

Para os 11 anos de microdados tratados sobre os docentes, temos a documentação:

- Um arquivo descrevendo a documentação que o acompanha “Leiamme”
- Um arquivo com o banco de dados (dataset) tratado e pseudoanonimizado (csv)
- Um arquivo com um dicionário sobre o tratamento/ padronização que foi dada ao banco (.ods)
- Um arquivo hash que contém um código de segurança para conferência da versão do Dataset (txt / md5)

Os dados serão coletados do banco de dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), referentes aos docentes das Instituições de Ensino Superior (IES), no período histórico de 2009 a 2019.

Documento em .txt com instruções sobre os arquivos que devem ser baixados para acessar o banco de dados e a informação das modificações realizadas pela equipe de pesquisadores responsáveis, bem como o fluxo dos processos de extração.

Os padrões de metadados seguirão a política de curadoria do Repositório de Dados da Unifesp.

Por tratar-se de dados retirados de um banco público as informações sensíveis já foram tratadas em acordo com a Lei Geral de Proteção de Dados, de forma a excluir as informações sensíveis.

Os dados produzidos a partir da ferramenta Pentaho Data Integration (PDI), ficarão disponíveis para acesso aberto, sendo obrigatório a citação de autoria e/ou a inclusão da referência que acompanha cada dado.

Os dados retirados do banco de dados do Inep já são de acesso público, sendo obrigatório a citação da referência.

O SoU_Ciência possui um computador dedicado ao armazenamento dos dados, além de possuir conta no GCP- Google Cloud Platform (Storage e BigQuery), para armazenamento na nuvem, como forma de backup.

Hoje, o acesso aos dados através da plataforma Pentaho é dos pesquisadores envolvidos, no futuro, para acesso ao banco, será necessária a adoção do preenchimento de um formulário de solicitação de acesso, em atendimento aos critérios da Plataforma.

Todos os dados podem ser compartilhados, sem nenhum tipo de impedimento legal e/ou restrição, além do gerenciamento de acesso feito pela plataforma Pentaho (através do Formulário de solicitação de acesso) e da citação de autoria.

Os dados serão disponibilizados no repositório de dados da Unifesp que possui um banco estruturado e com política de segurança de dados e backup.

Através do Repositório de Dados da Universidade Federal de São Paulo.

Não há restrição de acesso ou compartilhamento, apenas a obrigatoriedade da citação da autoria.

A responsável técnica Ecila Alves de Oliveira Migliori.

Não há custo de armazenamento e acesso. Os servidores de armazenamento são da Universidade Federal de São Paulo.

Planned Research Outputs

Dataset - "Banco de dados a partir da base do Censo da Educação Superior - Docentes "

A pesquisa, no que tange a construção do banco de dados, está finalizada e possibilita a qualquer pesquisador sua reprodução ou reutilização. Porém, novas pesquisas e atualizações seguem em andamento

Service - "Construção de um banco de dados a partir da base do Censo da Educação Superior - Docentes "

Banco de dados criado a partir das bases de dados do Inep, mais especificamente a base do Censo do Ensino Superior, dados Docentes, no período de 2009 a 2019.

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?
Banco de dados a partir da base do Censo da Educaç ...	Dataset	Unspecified	Open	None specified		None specified	None specified	No	No
Construção de um banco de dados a partir da base d ...	Service	Unspecified	Open	None specified		None specified	None specified	No	No