

---

## Plan Overview

*A Data Management Plan created using DMPonline*

**Title:** UMA PROPOSTA DE VOCABULÁRIO BILÍNGUE DE PESTICIDAS ORGANOFOSFORADOS POR MEIO DA LINGUÍSTICA DE CORPUS: UM FOCO NO TRATO DA VARIAÇÃO DENOMINATIVA

**Creator:** José Victor de Souza

**Principal Investigator:** José Victor de Souza

**Data Manager:** José Victor de Souza

**Project Administrator:** Paula Tavares Pinto

**Contributor:** Marcelo de Freitas Lima

**Affiliation:** Universidade Estadual Paulista Julio de Mesquita Filho

**Template:** DCC Template

**ORCID iD:** 0000-0002-4587-4073

### Project abstract:

Brazil is one of the largest agricultural producers in the world and, therefore, one of the largest consumers of pesticides. Thus, scientific research in the field of pesticide chemistry has become increasingly imperative in the country. However, with regard to the field's terminology, translation and terminological problems were found in scientific texts dealing with pesticides within the category of organophosphates. The problems revolve mainly around the lack of consensus on the translations of terms from English to Portuguese, which generates noticeable denominative variation. Thus, this project has the general objective of inventorying terms of pesticide chemistry in Portuguese and English, within the subarea of organophosphates, with the following specific objective: a) defining the microstructure of the entry for a specialized vocabulary of the area, taking into account its particularities, the denominative variations found as well as their possible causes. Using computerized tools, a specialized bilingual corpus will be expanded, from which terms will be extracted by calculating the keyness score of each lexical item. When variants are identified among the terms, they will be classified according to their type. Then, statistical indices will be applied and a specialist in the area consulted to define the most suitable variant to be included in the entry of the entry. Finally, based on data from the corpus and interviews with the expert, the possible causes of the denominative variations found will be defined. The information gathered will be registered in a terminological record that will result in a bilingual vocabulary.

**ID:** 81159

**Start date:** 01-03-2022

**End date:** 01-02-2023

**Last modified:** 08-05-2023

**Grant number / URL:** 2021/08830-9

**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 customise 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

# UMA PROPOSTA DE VOCABULÁRIO BILÍNGUE DE PESTICIDAS ORGANOFOSFORADOS POR MEIO DA LINGUÍSTICA DE CORPUS: UM FOCO NO TRATO DA VARIAÇÃO DENOMINATIVA

---

## Data Collection

### What data will you collect or create?

Os dados coletados serão o conteúdo textual de artigos científicos, teses e dissertações. O formato escolhido para armazenamento foi o de texto simples (.txt). Todos os dados coletados para a pesquisa são de acesso livre estão disponíveis gratuitamente na web.

Os dados criados a partir da análise serão fichas terminológicas, salvas no formato de .doc.

Todos os dados estarão disponíveis no computador pessoal do pesquisador principal e na nuvem para acesso da equipe de pesquisa, tendo um volume de por volta de 100 MB.

### How will the data be collected or created?

Os dados serão coletados da web manualmente ou por web scraping. Todos os dados coletados para a pesquisa estão disponíveis na web.

As fichas terminológicas serão elaboradas pelo pesquisador principal com base em trabalhos científicos publicados anteriormente.

## Documentation and Metadata

### What documentation and metadata will accompany the data?

Os dados criados encontrar-se-ão em anexo aos relatórios e à dissertação do beneficiário, podendo ser consultados por meio do repositório da Unesp uma vez publicados. Todos os dados podem ser lidos por softwares leitores de texto ou PDF (Word, LibreOffice, Adobe e afins).

## Ethics and Legal Compliance

### How will you manage any ethical issues?

Não há questões éticas envolvidas.

### **How will you manage copyright and Intellectual Property Rights (IPR) issues?**

Por questões de direitos autorais, o corpus compilado em sua totalidade não será disponibilizado para acesso livre. Ademais, durante a análise do corpus, os textos são analisados por fragmentos (linhas de concordância) e não são disponibilizados em sua totalidade.

### **Storage and Backup**

#### **How will the data be stored and backed up during the research?**

Todos os dados serão armazenados no computador pessoal do pesquisador principal com backup na nuvem do Sketch Engine. Não há data de validade.

#### **How will you manage access and security?**

Somente membros do grupo de pesquisa têm acesso aos dados por meio das credenciais do Sketch Engine.

### **Selection and Preservation**

#### **Which data are of long-term value and should be retained, shared, and/or preserved?**

Os dados gerados poderão ser utilizados para pesquisas futuras do pesquisador principal ou de seus colegas de pesquisa. Nenhum dado será destruído ou retido.

#### **What is the long-term preservation plan for the dataset?**

A dissertação com a análise dos dados ficará disponível no repositório da Unesp. Os dados em si ficarão preservados na nuvem. Não há geração de custos.

## **Data Sharing**

### **How will you share the data?**

Os dados só serão acessíveis por meio da análise do pesquisador principal. Nenhum usuário terá acesso aos dados brutos a não ser membros do grupo de pesquisa do pesquisador principal. O compartilhamento dos dados brutos é feito por meio do Sketch Engine ou por meio de outros recursos da nuvem (Gmail, Drive, etc.). O glossário será disponibilizado para livre acesso.

### **Are any restrictions on data sharing required?**

Não se aplica.

## **Responsibilities and Resources**

### **Who will be responsible for data management?**

O pesquisador principal e os orientadores serão responsáveis pelos dados.

### **What resources will you require to deliver your plan?**

Todos os recursos necessários são disponibilizados pelo instituto.