

# README

## A global dataset of pandemic- and epidemic-prone disease outbreaks

**Torres Munguía, J.A.; Badarau, F.C.; Díaz Pavez, L.R.; Martínez-Zarzoso, I.; & Wacker, Konstantin M.**

Here you can find a new dataset of infectious disease outbreaks collected from the Disease Outbreak News and the Coronavirus Dashboard produced by the World Health Organization (WHO).

**METHODOLOGICAL NOTE:** The unit of analysis in the database is an **outbreak**, defined as the occurrence of at least one case of a specific disease in a country -or territory- during a particular year. Therefore, a country -or territory- cannot have more than one outbreak of the same disease in the same year, although it may experience outbreaks of different diseases within the same year. A country can only have multiple outbreaks of the same disease if they occur in different years.

Potential applications of the data include combining it with other information sources to explore the relationship between disease outbreaks and factors such as geographical, environmental, globalization, and socioeconomic conditions.

**About the project** The dataset is the result of collaborative work by a team of researchers from the University of Göttingen, the University of Groningen, and the University of Bordeaux. The project was made possible through financial support from the ENLIGHT network, the German Academic Exchange Service (DAAD), and the Federal Ministry of Education and Research (BMBF) in Germany.

The findings of the first version of the dataset have been published in Springer Nature's Scientific Data. Read the paper by clicking [here!](#). Additionally, the data, metadata, and the code to replicate the first version of this dataset are publicly available on Figshare. You can download them by clicking [here!](#)

**IMPORTANT NOTE** From October 2024, this project is being updated by Dr. Juan Armando Torres Munguía. In case of questions, requests, or collaborations, you can contact me via [GitHub](#), [X](#), or [here](#).

The last version of the dataset was updated on 24/12/2025 and contains information on 3450 **outbreaks**, associated with 91 infectious diseases that occurred from 01/01/1996 to 24/12/2025 in 236 countries and territories worldwide.