

EGU24-19940, updated on 09 Aug 2024

<https://doi.org/10.5194/egusphere-egu24-19940>

EGU General Assembly 2024

© Author(s) 2024. This work is distributed under the Creative Commons Attribution 4.0 License.



## A global database of natural hazards impacts reported in the scientific literature

Taís Maria Nunes Carvalho<sup>1</sup>, Jakob Zscheischler<sup>2</sup>, Christian Kuhlicke<sup>2</sup>, and Mariana Madruga de Brito<sup>2</sup>

<sup>1</sup>Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Universität Leipzig (carvalho@informatik.uni-leipzig.de)

<sup>2</sup>Helmholtz Centre for Environmental Research, Leipzig, Germany

The increased frequency and magnitude of natural hazards might significantly increase social, economic, and health impacts on society in the next decades. Existing studies and databases of natural hazard impacts have several limitations, such as (1) a low level of detail on how people were affected; (2) an underestimation of the impacts; (3) a limited geographical range; and (4) a lack of information on the source of the data. However, scientific publications, reports, and handbooks compose a large data repository that can provide valuable and trustworthy information on natural hazards. We are building a **global database on the impacts of natural hazards** that have been documented since 1950 in the scientific literature. We mapped global research on **climatological, hydrological, and meteorological extremes**, such as heatwaves and floods. We retrieved over 40 thousand full-text open-access papers from ScienceDirect and Pubmed. Documents were coded according to (i) relevance: if the study describes impacts from a natural hazard, (ii) hazard class: single or multiple hazards, and (iii) event assessment: specific or multiple climate-related events. A randomly selected sample of the documents was manually labeled and a classification model was trained to classify the remaining papers. We further developed an annotation scheme for marking information on climate-related hazards in scientific publications, such as the date and location of hazard and their impacts. The inter-annotator agreement analysis shows the complexity of this task and the high annotation quality in our corpus. This work fills a critical gap in information extraction tasks within the natural hazards research domain, providing a robust foundation for future studies and analysis.