

# **Towards a systematic use of web-text data to support geospatial analysis of major natural disaster and crisis events – Evidence from the Ahrtal 2021 flooding, Germany**

Vanessa Rittlinger<sup>1</sup>, Sheikh Mastura Farzana<sup>2</sup>, Xuke Hu<sup>3</sup>, Hema Pandey<sup>3</sup>, Stefan Voigt<sup>1</sup>, Christian Geiß<sup>1</sup>, Hannes Taubenböck<sup>1</sup>

<sup>1</sup>*German Aerospace Center (DLR), Earth Observation Center, Oberpfaffenhofen, Germany*

<sup>2</sup>*German Aerospace Center (DLR), Institute for Software Technology, Cologne, Germany*

<sup>3</sup>*German Aerospace Center (DLR), Institute of Data Science, Jena, Germany*

During and after natural disaster or crisis events, availability of relevant situation assessments, possibly in near real-time, are essential for many kinds of disaster and crisis relief activities. In this context, web-text data can be a valuable source of information, as they are often available in a timely manner. However, the challenge lies in the extraction and automated aggregation of meaningful insights from this unstructured data. In this talk we present ongoing work regarding an approach to extract geospatial disaster management information from web-crawl data. A special focus is put on the identification of event-related web information, including thematic information on the development and impact of a disaster event as well as the analysis of its geospatial context. As a proof of concept, we reanalyze a flood event that occurred in the Ahrtal valley, Germany in June 2021, resulting in significant damage and human and economic loss.

Crawled Web-text data is used to reconstruct temporal and spatial aspects of this disaster event. A combination of different natural language processing methods is used to filter relevant web-text documents, extract, analyze, and visualize event related information. The processing steps are subdivided into steps focusing on location and thematic-based filtering approaches as well as topic detection methods to classify meaningful information into event related classes. Additionally, a geospatial component is included in the information filtering and generation process and used for geographic characterization of the disaster impact. In this step, we will employ advanced geoparsing approaches that leverage mid-sized large language models like Mistral (7B) and geographic knowledge from OpenStreetMap to accurately extract geospatial information from texts, including fine-grained locations, such as streets, houses, and points of interest.

Apart from the geospatial and flood related information we also attempt to assess the temporal dynamics of the event. In an outlook we describe how web-text derived disaster information may be combined with other geospatial data sources such as satellite imagery to improve situational understanding and assess spatio-temporal dynamics of major natural disaster or crisis events.