Comprehensive assessment of flood socioeconomic impacts through text-mining

Mariana Madruga de Brito[1], Jan Sodoge[1], Heidi Kreibich[2], and Christian Kuhlicke[3]

[1]Helmholtz-Centre for Environmental Research
[2]GFZ German Research Centre for Geosciences
[3]UFZ Helmholtz Centre for Environmental Research

March 05, 2024

Abstract

In July 2021, Germany experienced its costliest riverine floods in history, with over 189 fatalities and a staggering \euro33 billion in damages. Following this event, news outlets widely disseminated information on the flood's aftermath. Here, we demonstrate how newspaper data can be instrumental in the assessment of flood socioeconomic impacts often overlooked by conventional methods. Using natural language processing tools on 14,888 unique newspaper articles, we estimate the impacts of the 2021 flood on various sectors and critical infrastructure, including water contamination, mental health, and tourism. Our results revealed severe and lasting impacts in the Ahr Valley, even months after the event. At the same time, we identified smaller-scale yet widespread impacts across Germany, which are typically overlooked by existing impact databases. Our approach advances current research by systematically examining indirect and intangible flood impacts over large areas. This underscores the value of leveraging complementary text data to provide a more comprehensive picture of flood impacts.

Hosted file

`de_Brito_et_al_WRR.docx` available at https://authorea.com/users/634520/articles/721071-comprehensive-assessment-of-flood-socioeconomic-impacts-through-text-mining

Comprehensive assessment of flood socioeconomic impacts through text-mining

Mariana Madruga de Brito[1], Jan Sodoge[1,2], Heidi Kreibich[3], Christian Kuhlicke[1,2]

[1] Department of Urban and Environmental Sociology, UFZ-Helmholtz Centre for Environmental Research, 04318 Leipzig, Germany

[2] Institute of Environmental Science and Geography, University of Potsdam, 14476, Potsdam-Golm, Germany

[3] GFZ-German Research Centre for Geosciences, 14473 Potsdam, Germany

Corresponding author: Mariana Madruga de Brito (mariana.brito@ufz.de)

Key Points:

- Text-based impact assessments can provide a comprehensive picture of flood impacts at a large spatial scale.
- Our approach allows capturing a wider range of impact classes compared to conventional assessments (e.g. mental health, tourism, and water contamination)
- While the Ahrweiler region was the most affected during the 2021 floods in Germany, smaller-scale events also occurred throughout the country.

Abstract

In July 2021, Germany experienced its costliest riverine floods in history, with over 189 fatalities and a staggering €33 billion in damages. Following this event, news outlets widely disseminated information on the flood's aftermath. Here, we demonstrate how newspaper data can be instrumental in the assessment of flood socioeconomic impacts often overlooked by conventional methods. Using natural language processing tools on 14,888 unique newspaper articles, we estimate the impacts of the 2021 flood on various sectors and critical infrastructure, including water contamination, mental health, and tourism. Our results revealed severe and lasting impacts in the Ahr Valley, even months after the event. At the same time, we identified smaller-scale yet widespread impacts across Germany, which are typically overlooked by existing impact databases. Our approach advances current research by systematically examining indirect and intangible flood impacts over large areas. This underscores the value of leveraging complementary text data to provide a more comprehensive picture of flood impacts.

1 Introduction

Flood disasters are predicted to become more frequent in many regions due to climate, land use, and population demographic changes (Bruno Merz et al., 2021; Seneviratne et al., 2021), as well as the increasing interconnectedness of coupled human-water systems (Fekete et al., 2021; Yoon et al., 2021). Despite this expected increase, our knowledge of flood's direct and indirect impacts remains fragmented. Existing studies based on damage functions (Carrera et al., 2015; Merz et al., 2010) and surveys (Kreibich et al., 2010) often provide only a partial picture of flood consequences. A major challenge of these studies is that they tend to focus on direct impacts and analyze only a single impact type at a time (e.g. critical infrastructure - Qiang et al., 2020, agriculture losses - Chen et al., 2019; Rahman & Di, 2020; Tapia-Silva et al., 2011, damage to buildings - Gerl et al., 2014; Serpico et al., 2012, fatalities - Papagiannaki et al., 2022). Moreover, detailed damage estimation studies often have limited geographic scope. These limitations are linked to difficulties in obtaining reliable loss estimates for a wide range of impact types simultaneously due to the time-or data-intensive nature of conventional impact assessment methods (Carrera et al., 2015; Kreibich et al., 2010; Merz et al., 2010).

An additional significant gap in impact assessments refers to the emphasis on severe and large-scale floods. Nevertheless, 99.7% of all disasters worldwide between 1990 and 2013 were smaller events involving fewer than 30 deaths or less than 5,000 houses destroyed (UNISDR, 2015). Thousands of smaller-scale floods thus go unrecorded because they do not generate sufficiently high impacts at the national or international levels. However, they cause a continuous stream of local losses (UNDRR, 2021). As a result, flood impacts have historically been underestimated (Sanders et al., 2022).

The scarcity of datasets covering both direct and indirect impacts, as well as low and high-impact events, severely constrains our ability to understand the far-reaching consequences of floods. Hence, comprehensive assessments of how floods affect society can better support effective adaptation: Impact data of preferably diverse sectors (e.g. tourism, agriculture, infrastructure), including tangible and intangible losses, are needed to improve our understanding of impact dynamics (Kellermann et al., 2020) and serve as validation data for impact-based forecasting systems (Mitheu et al., 2023). Moreover, such assessments can equip decision-makers with a deeper understanding of all the potential flood consequences. Based on that, they can identify areas with severe damages (Kryvasheyeu et al., 2016) as well as those with low impacts but which face frequent 'nuisance' floods (Moftakhari et al., 2018).

As researchers face the need for comprehensive methods to produce impact datasets, the use of text data for extracting information on natural hazard consequences is flourishing (de Brito et al., 2020; Franceschini et al., 2022; Hao & Wang, 2020; L. Li et al., 2021; Liu & Jensen, 2018). Texts offer a rich yet untapped complement to the more structured data types traditionally used in research (Gentzkow et al., 2019). Texts are an essential component of our lives – it is how historical events are recorded, and citizens' concerns are documented. Within this context, computational text analysis offers a unique opportunity to automatically extract information on flood impacts stored in vast amounts of text.

Here, we use newspaper texts to infer the socioeconomic impacts of the July 2021 floods in Germany, particularly in the Ahr River region but also in the South and East of the country. The disaster in the Ahr River was Germany's most severe flood event ever recorded, with unprecedented impacts and unmatched precipitation values (Junghänel et al., 2021; Kreienkamp et al., 2021). It affected more than 40,000 people (Zeit Online, 2021), causing over 189 direct fatalities and at least €33 billion in damage (Dietze et al., 2022; Fekete & Sandholz, 2021; Kron et al., 2022; Thieken et al., 2022). The main flood event occurred between the 12th and 19th of July (Junghänel et al., 2021), with an estimated return period exceeding 500 years in the Ahr river (Kreienkamp et al., 2021). Given its widespread consequences and extensive media coverage, we view this disaster as a test case for investigating whether newspaper articles can provide reliable data on flood impacts for both large and small-scale events. By examining this extreme case, we can verify whether smaller floods happening far from the main disaster area are potentially "ignored" by the media or if it is possible to capture them based on local and regional news outlets.

Our study provides a comprehensive overview of the multifaceted effects of the 2021 floods in Germany, demonstrating that newspaper articles constitute a rich data source to assess flood socioeconomic impacts for both large and small-scale events. Using natural language processing (NLP) tools on a newspaper text corpus, we provide empirical insights into the impacts of the 2021 floods in Germany and their varying spatio-temporal magnitude. We examine indirect and intangible impacts often neglected by conventional methods. As such, our NLP-based model shifts the focus away from damages to buildings and fatalities to encompass 20 distinct socioeconomic impact classes (e.g. mental health, water contamination) across the country. We evaluate our text-based estimates of impacts against a series of independently collected data to examine if the media reporting is linked to observed water levels and to how the government perceived the flood consequences (i.e. by declaring a status of emergency).

2 Methods

2.1 Newspaper data collection and pre-processing

We used NLP tools to determine text patterns that predict the socioeconomic impacts of floods on multiple sectors based on newspaper data from a news aggregator database (www.wiso-net.de). To identify relevant articles, we searched the title and subtitle of the articles using the flood-related keywords in Box 1 in German. To remove news mentioning floods in other countries, we only considered articles with a country tag related to Germany. We tested different sets of diverse synonyms for "flood" and manually examined the outcomes to verify whether or not all relevant articles were found. Thus, following a methodology similar to Hase et al. (Hase et al., 2021) search keywords were optimised to find relevant articles while accepting a slightly lower precision, i.e. including a few irrelevant articles. To verify the accuracy of our search, we randomly selected a sample of 100 articles. Of these, all were related to flood events and 4 were specific to floods in other countries. These irrelevant articles were later excluded by our location extraction approach (see Section 2.5). All the 250 printed and online news outlets within the Wiso newspaper database were considered. These comprise 144 local, 63 regional, and 43 national outlets (see supplementary Table S1 for a list). Based on this search, we retrieved 26,113 articles published between 01.07.2021 and 31.12.2021 in daily and weekly news outlets, covering local, regional, and national news. An overview of the procedure used to create the multi-sector impact database is provided in Supplementary Fig. S1.

Box 1. Search keywords used in the title and subtitle of newspaper articles. A translation to English is also provided

German *(original)*: Hochwasser* OR Überschwemmung* OR Überflutung* OR "heftiger Regen" OR Sturzflut OR Starkreg* OR Flutkatastrop*

English *(translation)*: flood* OR inundation* OR flooding* OR “heavy rain” OR “flash flood” OR “heavy rain” OR “catastrophic flood*”

To ensure that our corpus consisted of unique articles, we computed Jaccard similarity (de Brito et al., 2021) scores between all article pairs. The Jaccard score measures the proportion of common word sequences between two texts (Mullen, 2016). After analyzing the articles’ similarity histogram, and manually reading a set of 30 article pairs, we removed duplicate articles with a similarity score greater than 0.75, where 1 indicates full similarity. As a result, 14,888 unique articles remained for further analysis (Fig. 1).

To clean the raw text data, we decomposed each newspaper article text into single words and converted them to their root form (i.e. lemmatization). We then removed special characters, stopwords (e.g. pronouns, articles), words that consisted of less than two characters, and words used less than ten times in the entire corpus.

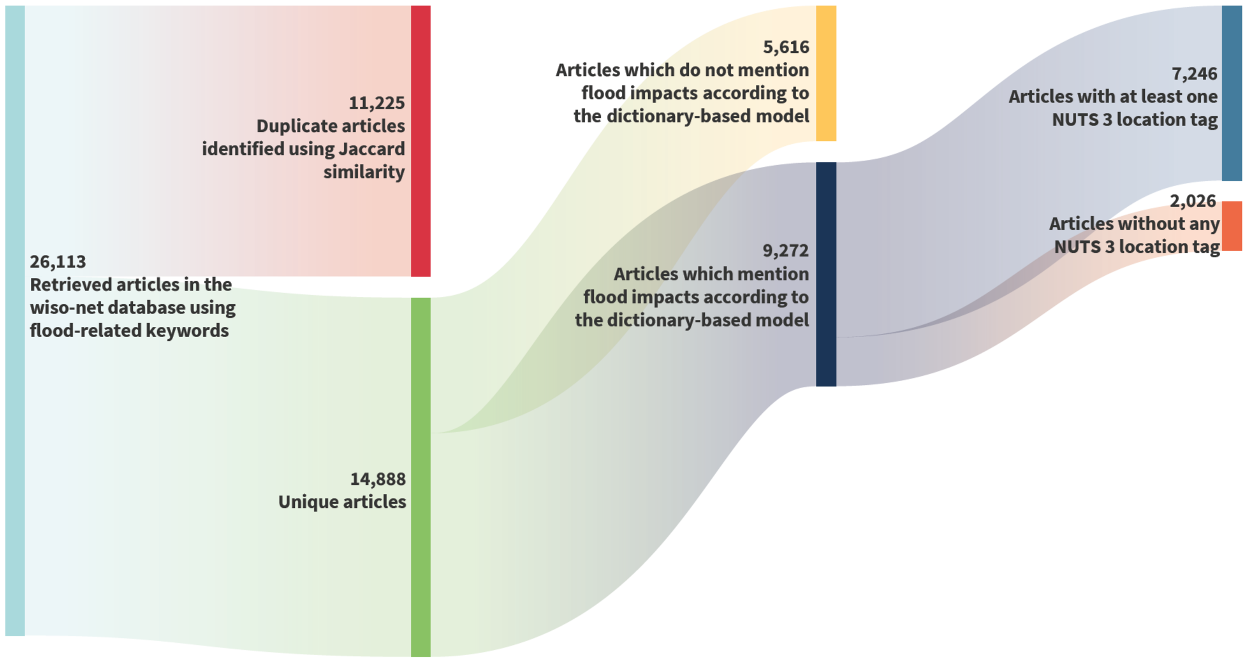


Fig. 1 Sankey plot with the steps of impact classification assessment procedure and the number of articles in each step. NUTS 3 (Nomenclature of Territorial Units for Statistics) comprises German districts (n = 401).

2.2 Impact typology

To classify the articles according to specific impacts, we developed a multi-sector flood socioeconomic impacts typology using both inductive and deductive approaches. For the inductive one, we built a topic model to identify dominant topics from the data itself. By doing so, we wanted to avoid missing relevant impacts not considered in previous studies. The Latent Dirichlet Allocation (LDA) algorithm was used, given its capability to identify latent topics in large datasets. LDA assumes that each document is a mixture of several topics and each topic is a distribution of words. The number of topics was set to 65 after computing the topic coherence for varying numbers of topics from 10 to 120. Topic coherence was selected as a measure to determine the optimal number of topics because it offers a quantitative assessment of the meaningfulness and organization of topics within a corpus. As an outcome of the LDA, a list of the most common words for each topic was obtained (Supplementary Table S2). This provided insight into the themes mentioned in the articles, such as economic losses and damages to transport.

While informative, the LDA results often overlooked specific impact types (e.g. tourism, water contamination, agriculture, and livestock). To address this, we also followed a knowledge-driven deductive approach, leveraging our expertise in flood impact assessment (Merz et al., 2010; Papagiannaki et al., 2022) and existing research (Hammond et al., 2015; Koks et al., 2022). As a result, we developed a comprehensive typology of 20 multi-sector flood socioeconomic impact classes (Supplementary Table S3).

2.3 Impact data annotation

After defining this typology, we manually read and annotated a random subset of the unique articles with these 20 specific impact categories. In order to build accurate machine learning classification models, we aimed to have at least 50 positive labels (e.g. articles reporting an impact) for each impact class. Since some articles contain more than one impact class, only 640 articles were read, totalizing 1,582 positive impact labels (see the number of labeled data per impact class in Supplementary Table S5). The articles were hand-coded by the first author with the support of the second author, following the principles of qualitative content analysis (de Brito et al., 2021; Schreier, 2012). This annotated impact dataset served as the basis for evaluating the classification models.

2.4 Classification models for the assessment of floods' socioeconomic impacts

We developed five multilabel classification models to assess flood's socioeconomic impacts from the unique articles (n=14,888). The term "multilabel" implies that each article may contain multiple impact classes. To construct these models, we considered both dictionary-based as well as machine-learning approaches.

To develop the dictionary-based algorithm, we relied on domain knowledge by selecting representative and comprehensive keywords. We also analyszed word frequencies, word co-occurrences, and topic modelling results. For each impact class, we considered a series of seed terms (e.g. energy, transportation) and used the *word2vec* algorithm in R to identify similar terms. This approach measures the semantic similarity between words and helps to expand the set of relevant terms. The resulting dictionary (Supplementary Table S4) was used to identify sentences where these keywords occurred and tag them as related to a given impact class. The resulting model predictions were evaluated against the manually annotated dataset (n=640 articles).

In addition to the dictionary-based approach, we also tested four machine learning models using the Python *scikit-multilearn* library: (i) logistic regression, (ii) support vector machine, (iii) random forest, and (iv) XGboost. A description of each of these models is provided in the *scikit-multilearn* library user guide (Szyma & Kajdanowicz, 2016). To train and test the models, we randomly split the annotated dataset (n=640) into 70% for training and 30% for testing.

We evaluated all five models against the manually annotated dataset. To do so, we computed performance metrics such as accuracy (Eq. 1), precision (Eq. 3), recall (Eq. 2), and F-score (Eq. 4).

$$Accuracy = \frac{True\ negative + True\ positive}{True\ positive + False\ positive + True\ negative + False\ negative} \quad \text{Eq. 1}$$

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \quad \text{Eq. 2}$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \quad \text{Eq. 3}$$

$$F\ score = \frac{Precision \times recall}{Precision + recall} \quad \text{Eq. 4}$$

2.5 Location extraction

To extract the geographical locations mentioned in each article, we followed the three-step method proposed by Sodoge et al.(Sodoge et al., 2022) (see Supplementary Fig. S2). First, we used the Google BERT named entity recognition (NER) algorithm (Devlin et al., 2018) in R to identify all location names mentioned in the text. These include names of regions, cities, neighbourhoods, streets, and rivers.

Next, we matched the identified locations with the GeoNames gazetteer (GeoNames, 2022) to obtain their corresponding geographical coordinates. Given that the names of these locations are not unambiguous in the GeoNames dataset, we excluded those that appeared more than five times in the matched data. This step was necessary to eliminate locations that cannot be unambiguously identified. For example, hundreds of streets in Germany are named "Hauptstraße" ("main street" in English). This measure affected only detailed location names (e.g., neighborhoods and streets) as district names are unique according to the European Union Nomenclature of Territorial Units for Statistics (NUTS).

Finally, to remove locations that were mentioned but not affected by floods (e.g. location of the newspaper headquarters), we excluded regions with exceptionally low density of location tags. For instance, newspaper headquarters tend to be mentioned only once at the beginning of the articles in the Wiso database. To this end, an agglomerative hierarchical clustering algorithm was used. This method treats each identified location as a single cluster and iteratively merges the two closest ones. The largest cluster was considered as the location of the flood impact, while any distant locations were treated as outliers and excluded from the analysis. For instance, if an article mentioned three districts, including two that were closely located (district A and B) and one far away (district C), our algorithm would only map the two closely located districts due to their high location tag density. We then aggregated the obtained data by their NUTS level 3, which refers to the most detailed NUTS classification (i.e. district scale).

The capabilities of the three-step location extraction method were evaluated in detail in previous research (Sodoge et al., 2022). To evaluate its performance in this study, we drew a random sample of 50 unique newspaper articles. By manually reading them, we assessed whether the extracted locations (n=236) were mentioned in the article. We found that all locations extracted (100%) were indeed mentioned in the text. However, not all referred to the location of the flood consequences. As such, an accuracy of 79% was obtained for estimating the location of the impact. Mismatches occurred due to articles mentioning nearby but unaffected locations (e.g. firefighters from 'city A' supported evacuating residents in 'city B').

2.6 Flood impact statements (FIS) spatiotemporal database

To identify the time of the impact occurrence, we considered the time stamps of the newspapers. The identified impacts with a corresponding location were integrated into a final spatial-temporal dataset of flood impact statements (FIS). A FIS consists of a sentence that describes a flood impact across 20 different classes (Supplementary Table S2) and locates it at the NUTS 3 scale. One article can contain multiple FIS of different classes.

2.7 Empirical evaluation of the FIS dataset

To verify to which extent the FIS dataset corresponds to observed reality, we evaluated it against: (i) the flood emergency levels ensued by the German Civil Defense, (ii) the number of flood fatalities recorded in each district, (iii) the maximum flood water levels observed in each district in July 2021, and (iv) a flood damage model (Table 1).

Standard non-parametric tests were used to measure the strength and direction of the relationship between the FIS impact dataset and these independent variables. We computed Spearman rho coefficients between the data in Table 1 and the FIS impact dataset. Furthermore, the Kruskal-Wallis test was performed to test the differences in distribution between the number of FIS and categorical variables (i.e. flood emergency levels and maximum flood water level observed).

To verify biases introduced by the media reporting as more news are produced in heavily populated cities, we contrasted the FIS magnitude with the population size in each district (Supplementary Fig. S5). To this end, we computed the coefficient of determination and the Spearman rho coefficients. All analyses were performed in R (version 4.1.3), all statistical tests were two-sided, and statistical significance was considered when *p*-value <0.05.

It should be highlighted that the evaluation was restricted due to the lack of available data. For some FIS classes, a comparison was not feasible as no other data-driven assessment exists for the entire country (e.g. mental health, physical injuries). For instance, damage models for the 2021 event have been produced for only six districts despite many others being affected. Also, some of the existing impact data have temporal limitations. Indeed, the water gauge levels dataset provided by LHP(LHP, n.d.) covered only the month of July 2021.

Table 1. Overview of datasets used for evaluation, their spatial and temporal scale, data sources, and impact classes used for correlation analysis

Data	Spatial scale	Temporal scale	Source	Comparison performed with
Flood emergency levels of each district according to 3 classes: not affected, affected, and catastrophe	NUTS 3	July 2021	BBK(BBK, 2021)	Each of the individual 20 FIS classes and the sum of all FIS
Number of fatalities	NUTS 3	July 2021	Thieken et al.(Thieken et al., 2022)	Loss of life FIS
Maximum flood water level observed according to 5 classes: no flood, minor, moderate, major and extreme flood	Point data	July 2021	LHP(LHP, n.d.)	Each of the individual 20 FIS classes and the sum of all FIS
Flood damage model for buildings	NUTS 3*	July 2021	Sieg and Thieken(Sieg & Thieken, 2022)	Damage to buildings FIS
Population size of each district	NUTS 3	December 2021	Federal Statistical Office(Federal Statistical Office, 2022)	Sum of all FIS

*The model was produced only for six districts located in the federal states of North Rhine-Westphalia and Rhineland-Palatine: Ahrweiler Eifelkreis Bitburg-Prüm, Euskirchen, Heinsberg, Rhein-Erft-Kreis, Städteregion

Aachen. The final model results are not yet published, and the linked reference refers to previous studies using the same model.

3 Results

3.1 Consequences of the 2021 German floods on multiple sectors according to newspaper articles

Following the devastating floods in Germany in July 2021 (Junghänel et al., 2021; Kreienkamp et al., 2021), both printed and online media provided extensive coverage of their aftermath. Over 26,000 flood-related news were published between July and December 2021 in Germany, which became the primary data source for this study.

To uncover and categorize the consequences of this disaster as reported by newspapers, we designed five different NLP-based classification models. After evaluating their performance against the manually annodated dataset, we found that the dictionary-based approach was the most accurate, with an average accuracy of 98% and an F-score of 91% (Supplementary Fig. S3 and Supplementary Table S5). As such, we used it throughout this study to assign one or more impact classes to each article. By applying the dictionary-based model to the newspaper corpus, we identified 20,052 flood impact statements (FIS). Each FIS consists of a sentence that describes and locates a flood impact at the district level.

Our analyses revealed that the floods had far-reaching consequences on multiple sectors, affecting critical infrastructure, individuals, buildings, and different economic sectors (Fig. 2). Numerous critical infrastructures were disrupted, including roads, wastewater treatment plants, water supply, mobile phone services, and electricity, gas, and heating systems. Media reports on ‘transport and mobility’ FIS were particularly prevalent (n=3,038), describing disruptions to roads, railways, and the collapse of bridges. Of the 112 bridges located in the flooded area of the Ahr valley, 62 were destroyed(Schäfer et al., 2021). Until July 2022, none of the 18 bridges on the Ahr river were functional, and temporary crossings are currently being used (The Guardian, 2022). Furthermore, major access routes, 600 km of railway, and 50 railway bridges were damaged (Schäfer et al., 2021), causing significant disruptions to transportation.

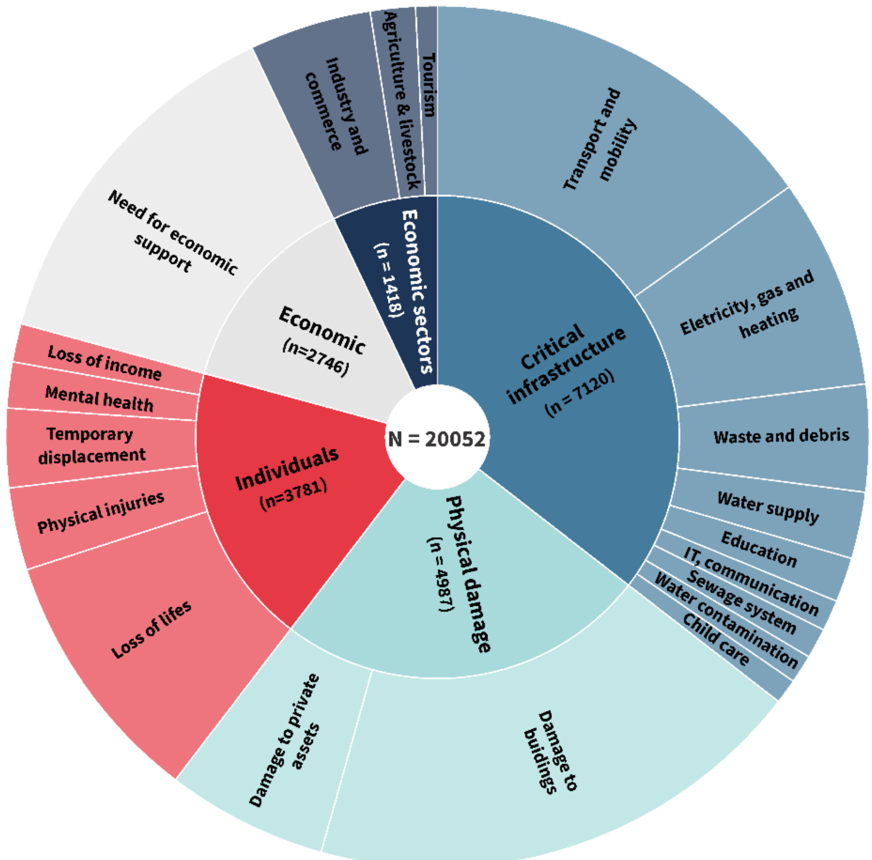


Fig. 2 Number of flood impact statements (FIS) according to different socioeconomic impact classes. See Supplementary Table S2 for a description of each FIS class.

Besides impacts on transport, many articles reported on ‘electricity, gas and heating’ impacts (n=1,580). More than 200,000 households were left without energy - some for months (Reuters, 2021). The contamination of rivers and drinking water due to floating and bursting oil tanks (Bosseler et al., 2021; Koks et al., 2021) was also a major concern, as reported in 214 articles. FIS on ‘physical damage to buildings’ (n=3,777) were also predominant. These statements describe impacts such as flooded cellars and damaged commercial buildings. In the states of North-Rhine Westphalia and Rhineland Palatinate alone, at least 200,000 buildings were affected, with insurance losses on residential buildings and household goods totaling €6.5 billion(GDV, 2021). Even a year later, many buildings and infrastructures have not been rebuilt (Nick et al., 2023).

Impacts on individuals were primarily associated with reports on ‘loss of life’ (n=1,933). FIS regarding fatalities were frequently linked with ‘damage to building’ and ‘transport and mobility’ FIS (Supplementary Fig. S4), which can provide insight into the circumstances surrounding these incidents. FIS on ‘mental health (n=345) mainly involved articles that describe how people suffered adverse effects such as post-traumatic stress disorder. These articles discussed the availability, or the lack thereof, of post-disaster support and mental health care services.

FIS on economic sectors reported mainly impacts on ‘industry and commerce’ activities (n=917). Despite the importance of tourism in the Ahr valley (Vaughan, 2022), the FIS pertaining to this sector (n=163) were less prominently described in the media.

Overall, multiple FIS were reported concomitantly. In fact, of all unique articles mentioning an impact and their location (n=7,246), 55.1% mentioned two or more FIS of different classes simultaneously. In particular, articles that referred to impacts on 'education' FIS also reported on 'child care' problems (Spearman rho test = 0.34, 0.33-0.36 95% confidence interval - CI, $p <$ 0.0001). Moreover, infrastructure-related impacts (e.g. 'water supply' and 'eletricity, gas and heating') were often reported simultaneously (Supplementary Fig. S4). These linkages shed light on sectors that would benefit from joint adaptation measures.

3.2 Magnitude of flood socioeconomic impacts

To estimate the magnitude of the socioeconomic impacts, we aggregated the counts of FIS over time (01 July to 31 December 2021). This choice was motivated by the observation that disasters with widespread effects typically attract heightened media coverage, as shown by previous studies (Kryvasheyeu et al., 2016; Z. Li et al., 2018; Sodoge et al., 2023). The FIS magnitude was then compared with a series of independent data to investigate the reliability of this assumption.

Overall, we found that trends in the socioeconomic impacts of floods can be robustly inferred from the FIS magnitude derived based on news coverage. Fig. 3a shows that the impact magnitude increases with the proximity to stream gauges with recorded extreme floods (Fig. 3b). Districts with 'extreme', and 'major' flood levels have significantly more FIS in July (Kruskal-Wallis test, $p < 0.0001$) (Fig. 4a). The districts with the highest FIS magnitude are located along the Ahr river, where water levels reached their highest values since measurements commenced (Szymczak et al., 2022). In addition, FIS hotspots were identified in the south and east of Germany. In these regions, stream gauge measurements recorded extreme or major flood levels. It should be highlighted that the exact water levels are unclear in some gauges along the Ahr River because the flood wave damaged or destroyed many of the gauging stations (Ludwig et al., 2022; Szymczak et al., 2022).

The FIS magnitude (Fig. 3a) is also associated with the emergency levels declared by the German Civil Defense (Supplementary Fig. S5). Districts reported as having a 'flood catastrophe' presented a significantly higher number of FIS (Kruskal-Wallis test, $p < 0.0001$) (Fig 4b). Conversely, districts deemed as 'not affected' by floods had fewer FIS. The few outliers include the cities of Koblenz, Cologne, and Bonn, located on the Rhine river, a distributary of the Ahr river. Even though the Civil Defense considered them as not-affected regions, minor and moderate floods were recorded at the stream gauges of these cities (Fig. 3b). Indeed, several critical infrastructure, roads, and highways were disrupted in Koblenz, Bonn, and Cologne (Fekete & Sandholz, 2021; Kron et al., 2022; Schauff & Schneider, 2021). Since these cities are densely populated, the flood waters impacted many people (i.e. higher exposure), which explains the high number of FIS.

The correlation between the 'loss of life' FIS and the observed fatalities(Wikipedia, 2022) (Spearman rho test= 0.89, 95% CI 0.87-0.91, $p < .0001$) is very strong (Fig. 5a). The cities with the highest FIS magnitude also had the highest number of recorded fatalities: Ahrweiler and Euskichen, with 133 and 27 fatal victims respectively.

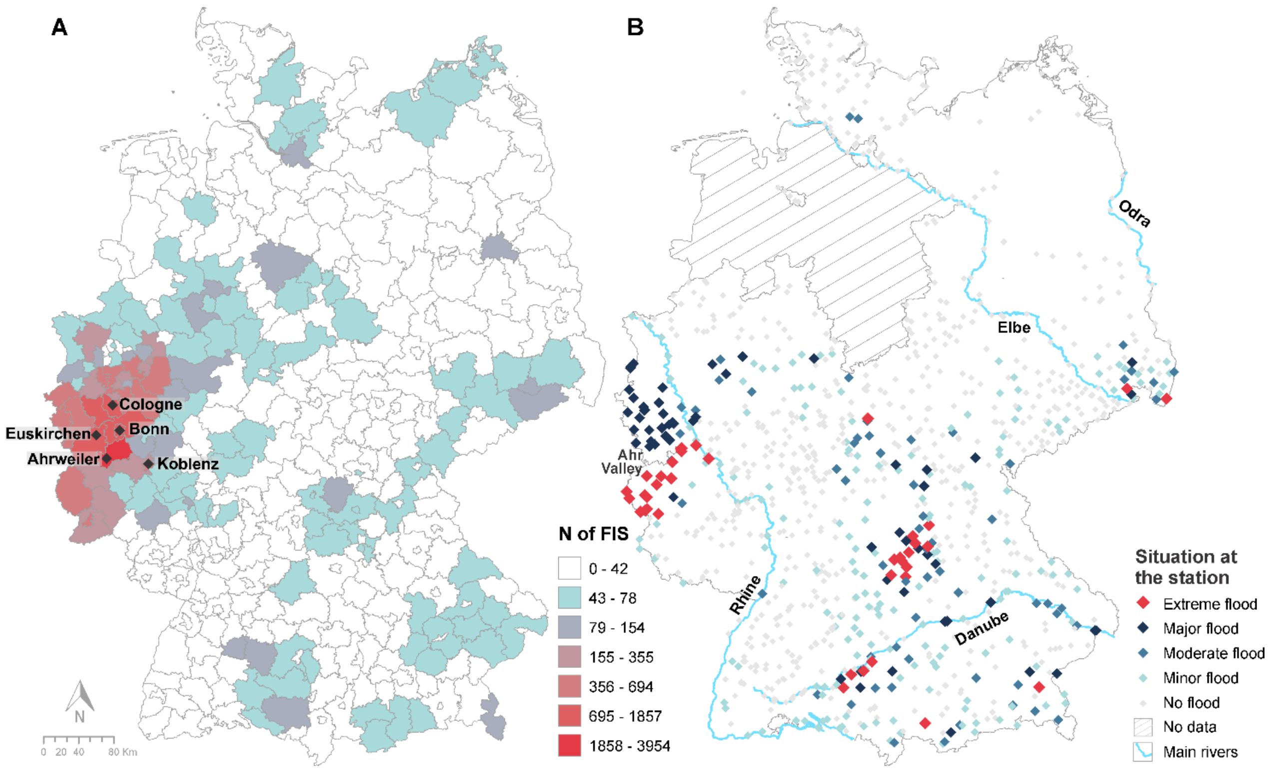


Fig. 3 Spatial distribution of flood events and their consequences. a, the sum of all FIS from 01 July to 31 December 2021. The districts with the highest FIS magnitude are highlighted. b, the maximum recorded flood levels at stream gauges between 01 and 31 July 2021. No stream gauge data is available for the Lower Saxony (NI).

We found strong and positive correlations between our data and modelled damages(Sieg & Thieken, 2022). For instance, ‘damage to buildings’ FIS correlate with modelled economic damages to buildings (Spearman rho = 0.98, 95% CI 0.80-0.99, p = 0.0007) (Fig. 5b) and the estimated number of affected buildings (Spearman rho test = 0.98, 95% CI 0.80-0.99, p < 0.0001) (Fig. 5c). Furthermore, moderate correlations were found between ‘industry & commerce’ FIS and modelled economic damages to industry and commerce (Spearman rho = 0.91, 95% CI 0.60-0.99, p = 0.0310). This suggests that our approach reflects patterns created by data-intensive damage models, with the advantage that our data covers large areas.

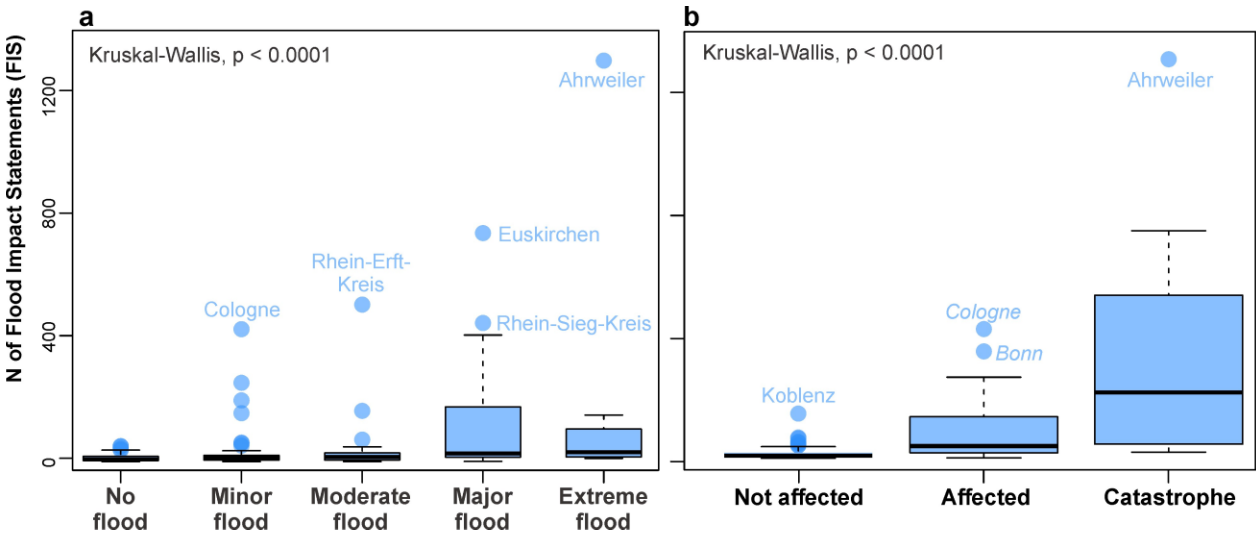


Fig. 4 FIS magnitude in July 2021 according to a, the maximum recorded flood levels at stream gauges and b, the German Civil Defence flood level.

An analysis of biases introduced by unequal media reporting in highly-populated cities suggests that the population size in each district does not correlate with the FIS magnitude (Supplementary Fig. S6). The coefficient of determination was R^2=0.10, indicating that the population size alone is not a good predictor of the FIS number, attesting to the robustness of our findings. In fact, districts with a relatively low population, such as Euskirchen and Ahrweiler, had a high FIS magnitude.

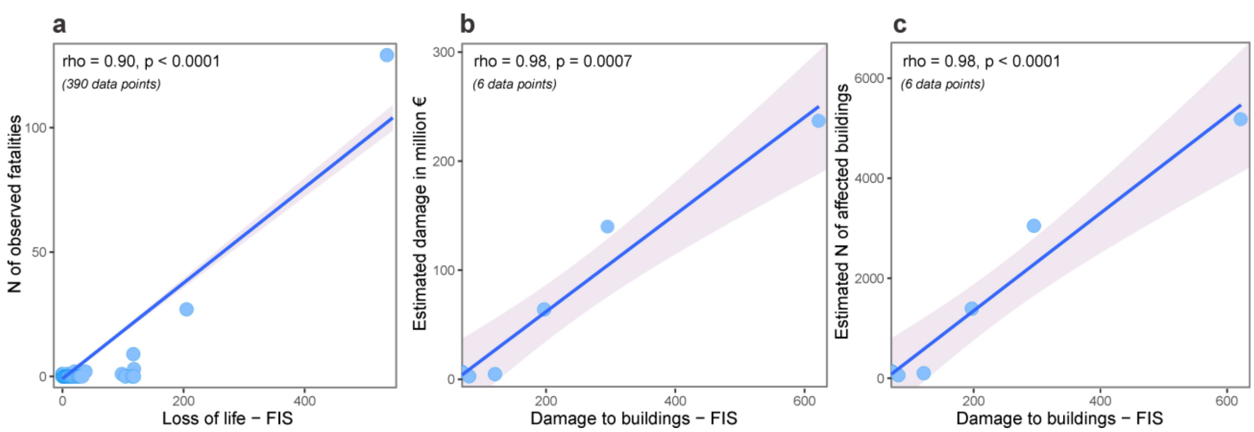


Fig. 5 Spatial correlation between different FIS classes and the a, number of observed fatalities. b, estimated damage in a million euros, c, estimated number of affected buildings. Each circle represents a district . The Spearman's rank correlation coefficient (rho) was used to examine the relationship between the FIS values (x-axis) and external data.

3.3 Dynamics of FIS across time, regions, and sectors

When analyzing the time transect of the FIS (Fig. 6a), we find that their peak was on 24 July, days after the main floods on 12-19 July 2021. This delay was expected, given the time required for news outlets to gather information and send reporters to flood-affected areas. The weekly number of FIS reduced by one order of magnitude three months later. While the FIS magnitude faded out in Bayern (BY) and Baden-Württemberg (BW), it persisted for the extreme events in North-Rhine Westphalia (NW) and Rheinland-Pfalz (RP) (Fig. 6b). The FIS magnitude decreased for most impact classes over time. However, its share remained high months after the floods for some impact classes (Supplementary Fig. S7).

In early July 2021 (Fig. 7a), minor floods occurred in BY, NW, BW, and Saxony-Anhalt (SH). These events are often overlooked in studies investigating the 2021 floods in Germany as they were not as destructive as the Ahr Valley event. Yet, they caused local disruptions and should not be dismissed. For instance, 53 fire brigades and more than 650 emergency services were called due to flooding in the east of BY from 05.07.2021 to 07.07.2021. This event led to one fatality, damaged buildings, and affected agricultural fields. As a result, 49 articles reported flood impacts in BY during the first ten days of July 2021. These findings underscore the importance of using complementary text-based approaches to assess flood impacts, as they can account for small flood events that are typically ignored in impact datasets such as HANZE (Paprotny et al., 2018) and EM-DAT (EM-Dat, 2021).

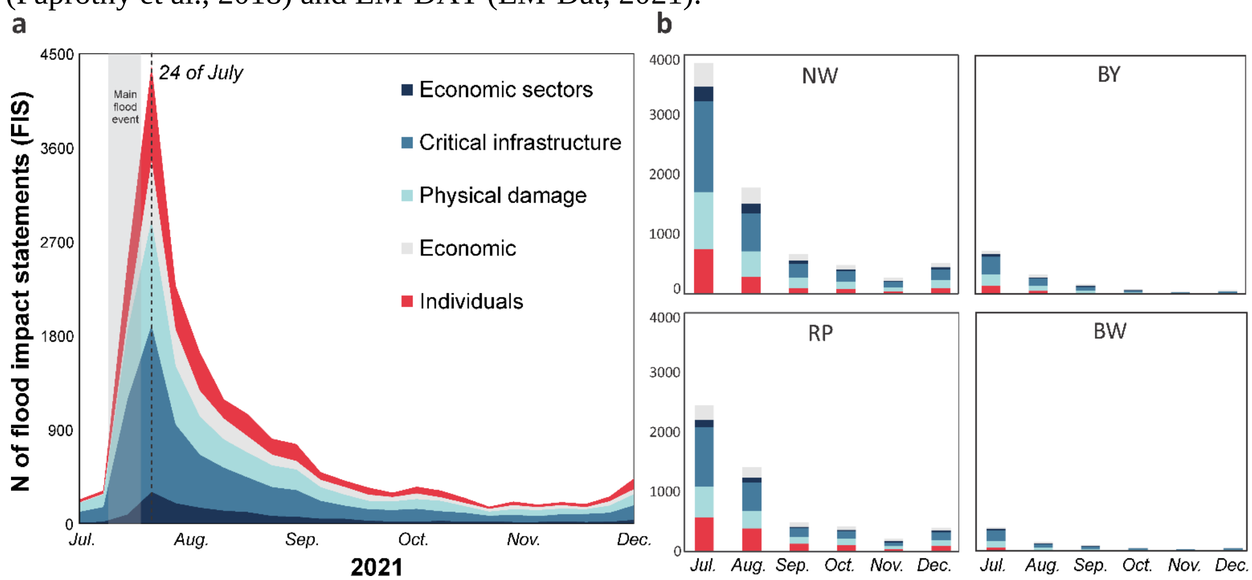


Fig. 6 Distribution of FIS over time and across different impact classes. a, aggregated FIS per week for Germany. b,aggregated FIS per month for the states of Nordrhein-Westfalen (NW); Rheinland-Pfalz (RP), Bayern (BY) and Baden-Württemberg (BW). Only states with the highest number of FIS are shown. The share of FIS for each of the 20 classes is given in Supplementary Fig. S7.

During and directly after the main flood event (Fig. 7b), multiple FIS were reported across Germany. FIS mentioning damage to buildings were rampant in NW (n=521), and RP (n=288), which presented record precipitation sums. Extreme flood levels were observed in SN, BW, and BY. However, the FIS magnitude was considerably lower in these states when compared to NW and RP. Indeed, only 199 FIS were identified in BY between 15 and 24.07.2021, whereas there were 1,070 FIS in NW and 524 in RP. The FIS magnitude drastically reduced two months after the floods (Fig. 7c). Nonetheless, it remained high for some impact classes, such as 'need for economic support', 'transport and mobility', and 'damage to buildings' in NW and RP (Supplementary Fig. S7). This trend may indicate that the floods had long-term effects on these sectors. As Koks et al. (2022) highlighted, until December 2021, roads, railways, education facilities, and wastewater treatment plants had not yet been reestablished in the Ahr Valley.

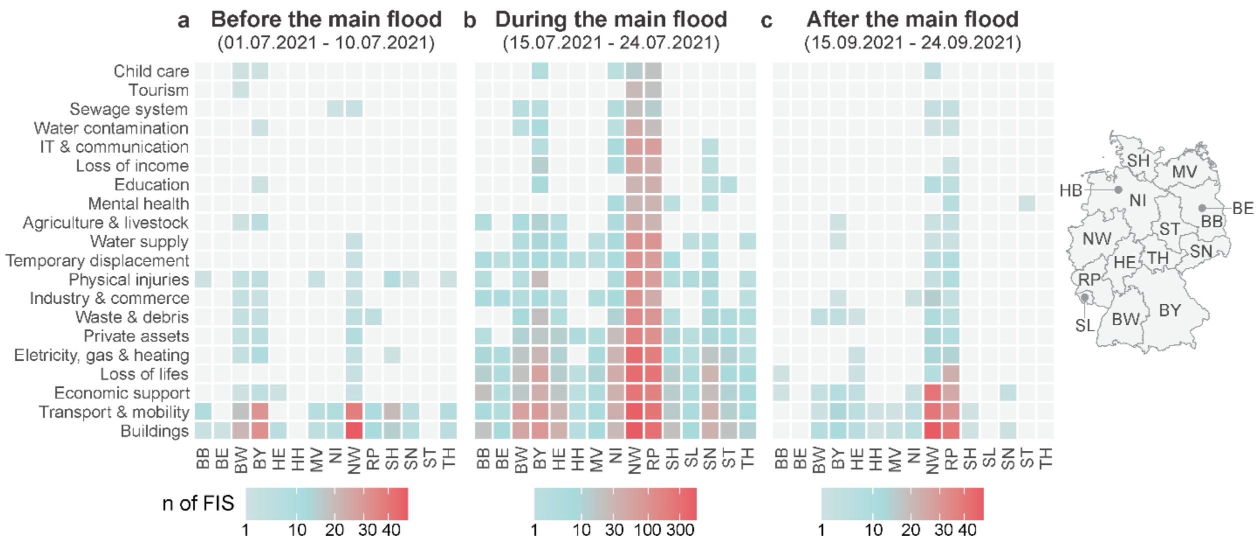


Fig. 7 Spatiotemporal evolution of FIS. a, before, b, during and directly after, c, after the main flood event. Logarithmic scales are used to represent the number of FIS. BW = Baden-Württemberg; BY = Bayern; BE = Berlin; BB = Brandenburg; HB = Bremen; HH = Hamburg; HE = Hessen; MV = Mecklenburg-Vorpommern; NI = Niedersachsen; NW = Nordrhein-Westfalen; RP = Rheinland-Pfalz; SL = Saarland; SN = Sachsen; ST = Sachsen-Anhalt; SH = Schleswig-Holstein; TH = Thüringen.

4 Conclusions

In this study, we assessed the magnitude of the socioeconomic impacts of the 2021 floods in Germany using a large corpus of newspaper articles published up to 6 months after the main flood events. Our empirical results revealed that the floods had far-reaching impacts across multiple sectors, underscoring the complex interplay between socioeconomic and infrastructure systems. The bulk of impacts results largely from ‘damage to buildings’ and ‘transport and mobility’ FIS, which triggered cascading effects in other sectors and infrastructure.

Our estimates show evidence for enhanced reporting of FIS in Ahrweiler, Euskirchen, and surrounding districts. In these regions, the FIS magnitude remained high even months after the event, which may indicate the lasting impacts of floods. Notably, we also detected small-scale events with fewer FIS in the south and east of Germany. These events were overlooked in previous studies and often go unnoticed in the national media or official disaster reports (Junghänel et al., 2021) and databases such as HANZE and EM-DAT. Despite their low impact, their cumulative effects over the years can be severe due to their high frequency (Kuhlicke et al., 2020). As such, despite the heightened attention to the Ahr Valley event, we could also detect small floods that took place in a similar time period.

Our findings support the hypothesis that the spatial and temporal trends in the occurrence of socioeconomic impacts of floods can be accurately inferred from news data. The FIS magnitude was associated with their proximity to gauges with extreme flood levels and correlated closely to official impact data (e.g. the number of fatalities) and modelled damages (e.g. economic losses to buildings). However, our approach has the added advantage of capturing a wider range of impact classes compared to conventional damage assessments. Indeed, text-based analyses such as the one presented here are one of the few means available for assessing impact classes such as mental health, tourism, and water contamination over large areas. Hence, our results demonstrate that text-based impact assessments can provide a comprehensive picture of flood impacts at a large spatial scale, going beyond the established "snap-shot" impact assessment approaches.

Acknowledgments

This work was conducted in the frame of the KAHR project (Climate Adaptation, Flood, Resilience) - funded by the German Federal Ministry of Education and Research (BMBF). This work was also supported by the Helmholtz Visiting Research Grant by The Helmholtz Information & Data Science Academy (HIDA).

Open Research

The datasets used for validation are available from the sources shown in Table 1. The resulting socioeconomic FIS database is provided in the supplementary material (SM1). The newspaper text data used in this analysis are restricted from public use due to copyright restrictions imposed by wiso-net.de.

References

BBK. (2021). *Zwischenbericht zur Flutkatastrophe 2021: Katastrophenhilfe, Soforthilfen und Wiederaufbau*.

Bosseler, B., Salomon, M., Schlüter, M., & Rubinato, M. (2021). Living with Urban Flooding: A Continuous Learning Process for Local Municipalities and Lessons Learnt from the 2021 Events in Germany. *Water*, *13*(19), 2769. https://doi.org/10.3390/w13192769

de Brito, M. M., Kuhlicke, C., & Marx, A. (2020). Near-real-time drought impact assessment: A text mining approach on the 2018/19 drought in Germany. *Environmental Research Letters*. https://doi.org/10.1088/1748-9326/aba4ca

de Brito, M. M., Otto, D., & Kuhlicke, C. (2021). Tracking Topics and Frames Regarding Sustainability Transformations during the Onset of the COVID-19 Crisis. *Sustainability*, *13*(19), 11095. https://doi.org/10.3390/su131911095

Carrera, L., Standardi, G., Bosello, F., & Mysiak, J. (2015). Assessing direct and indirect economic impacts of a flood event through the integration of spatial and computable general equilibrium modelling. *Environmental Modelling & Software*, *63*, 109–122. https://doi.org/10.1016/j.envsoft.2014.09.016

Chen, H., Liang, Q., Liang, Z., Liu, Y., & Xie, S. (2019). Remote-sensing disturbance detection index to identify spatio-temporal varying flood impact on crop production. *Agricultural and Forest Meteorology*, *269–270*, 180–191. https://doi.org/10.1016/j.agrformet.2019.02.002

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Retrieved from http://arxiv.org/abs/1810.04805

Dietze, M., Bell, R., Ozturk, U., Cook, K. L., Andermann, C., Beer, A. R., et al. (2022). More than heavy rain turning into fast-flowing water – a landscape perspective on the 2021 Eifel floods. *Natural Hazards and Earth System Sciences*, *22*(6), 1845–1856. https://doi.org/10.5194/nhess-22-1845-2022

EM-Dat. (2021). The Emergency Events Database.

Federal Statistical Office. (2022). Population statistics. Wiesbaden: Destatis. Retrieved from https://www.destatis.de

Fekete, A., & Sandholz, S. (2021). Here Comes the Flood, but Not Failure? Lessons to Learn after the Heavy Rain and Pluvial Floods in Germany 2021. *Water*, *13*(21), 3016. https://doi.org/10.3390/w13213016

Fekete, A., Aslam, A. B., de Brito, M. M., Dominguez, I., Fernando, N., Illing, C. J., et al. (2021). Increasing flood risk awareness and warning readiness by participation – But who understands what under 'participation'? *International Journal of Disaster Risk Reduction*, *57*, 102157. https://doi.org/10.1016/j.ijdrr.2021.102157

Franceschini, R., Rosi, A., del Soldato, M., Catani, F., & Casagli, N. (2022). Integrating multiple information sources for landslide hazard assessment: the case of Italy. *Scientific Reports*, *12*(1), 20724. https://doi.org/10.1038/s41598-022-23577-z

GDV. (2021). Ver-si-che-rungs-schä-den durch Flut-ka-ta-stro-phe bei rund sie-ben Mil-li-ar-den Euro. Retrieved from https://www.gdv.de/de/medien/aktuell/versicherungsschaeden-durch-flutkatastrophe-bei-rund-sieben-milliarden-euro-69800

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as Data. *Journal of Economic Literature*, *57*(3), 535–574. https://doi.org/10.1257/jel.20181020

GeoNames. (2022). GeoNames. Retrieved from https://www.geonames.org/

Gerl, T., Bochow, M., & Kreibich, H. (2014). Flood Damage Modeling on the Basis of Urban Structure Mapping Using High-Resolution Remote Sensing Data. *Water*, *6*(8), 2367–2393. https://doi.org/10.3390/w6082367

Hammond, M. J., Chen, A. S., Djordjević, S., Butler, D., & Mark, O. (2015). Urban flood impact assessment: A state-of-the-art review. *Urban Water Journal*, *12*(1), 14–29. https://doi.org/10.1080/1573062X.2013.857421

Hao, H., & Wang, Y. (2020). Leveraging multimodal social media data for rapid disaster damage assessment. *International Journal of Disaster Risk Reduction*, *51*, 101760. https://doi.org/10.1016/j.ijdrr.2020.101760

Hase, V., Mahl, D., Schäfer, M. S., & Keller, T. R. (2021). Climate change in news media across the globe: An automated analysis of issue attention and themes in climate change coverage in 10 countries (2006–2018). *Global Environmental Change*, *70*, 102353. https://doi.org/10.1016/j.gloenvcha.2021.102353

Junghänel, T., Bissolli, P., Daßler, J., Fleckenstein, R., Imbery, F., Janssen, W., et al. (2021). *Hydro-klimatologische Einordnung der Stark- und Zusammenhang mit dem Tiefdruckgebiet „Bernd" Dauerniederschläge in Teilen Deutschlands im vom 12. bis 19. Juli 2021*. *Deutscher Wetterdienst*. Retrieved from https://www.dwd.de/DE/leistungen/besondereereignisse/niederschlag/20210721_bericht_starkniederschlaege_tief_bernd.pdf?__blob=publicationFile&v=6

Kellermann, P., Schröter, K., Thieken, A. H., Haubrock, S.-N., & Kreibich, H. (2020). The object-specific flood damage database HOWAS 21. *Natural Hazards and Earth System Sciences*, *20*(9), 2503–2519. https://doi.org/10.5194/nhess-20-2503-2020

Koks, E., Van Ginkel, K., Van Marle, M., & Lemnitzer, A. (2021). Brief Communication: Critical Infrastructure impacts of the 2021 mid-July western European flood event. *Natural Hazards and Earth System Sciences Discussions [Preprint, in Review]*, (December), 1–11. Retrieved from https://doi.org/10.5194/nhess-2021-394

Koks, E. E., van Ginkel, K. C. H., van Marle, M. J. E., & Lemnitzer, A. (2022). Brief communication: Critical infrastructure impacts of the 2021 mid-July western European flood event. *Natural Hazards and Earth System Sciences*, *22*(12), 3831–3838. https://doi.org/10.5194/nhess-22-3831-2022

Kreibich, H., Seifert, I., Merz, B., & Thieken, A. H. (2010). Development of FLEMOcs – a new model for the estimation of flood losses in the commercial sector. *Hydrological Sciences Journal*, *55*(8), 1302–1314. https://doi.org/10.1080/02626667.2010.529815

Kreienkamp, F., Philip, S. Y., Tradowsky, J. S., Kew, S. F., Lorenz, P., Arrighi, J., et al. (2021). *Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021*. *Royal Netherlands Meteorological Institute (KNMI)* (Vol. 13). Retrieved from https://www.meteo.be/fr/infos/actualite/ce-que-lon-sait-sur-les-pluies-

Kron, W., Bell, R., Thiebes, B., & Thieken, A. (2022). The July 2021 flood disaster in Germany. In *HELP Global Report on Water and Disasters 2022*.

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. *Science Advances*, *2*(3). https://doi.org/10.1126/sciadv.1500779

Kuhlicke, C., Masson, T., Kienzler, S., Sieg, T., Thieken, A. H., & Kreibich, H. (2020). Multiple Flood Experiences and Social Resilience: Findings from Three Surveys on Households and Companies Exposed to the 2013 Flood in Germany. *Weather, Climate, and Society*, *12*(1), 63–88. https://doi.org/10.1175/WCAS-D-18-0069.1

LHP. (n.d.). Hochwasserzentralen. Retrieved from https://www.hochwasserzentralen.de/

Li, L., Bensi, M., Cui, Q., Baecher, G. B., & Huang, Y. (2021). Social media crowdsourcing for rapid damage assessment following a sudden-onset natural hazard event. *International Journal of Information Management*, *60*, 102378. https://doi.org/10.1016/j.ijinfomgt.2021.102378

Li, Z., Wang, C., Emrich, C. T., & Guo, D. (2018). A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods. *Cartography and Geographic Information Science*, *45*(2), 97–110. https://doi.org/10.1080/15230406.2016.1271356

Liu, L., & Jensen, M. B. (2018). Green infrastructure for sustainable urban water management: Practices of five forerunner cities. *Cities*, *74*(February 2017), 126–133. https://doi.org/10.1016/j.cities.2017.11.013

Ludwig, P., Ehmele, F., Franca, M. J., Mohr, S., Caldas-Alvarez, A., Daniell, J. E., et al. (2022). A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe. Part 2: Historical context and relation to climate change. *NHESS Discussions*, (July 2021), 1–42. Retrieved from https://doi.org/10.5194/nhess-2022-225
Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Assessment of economic flood damage. *Natural Hazards and Earth System Sciences*, *10*(8), 1697–1724. https://doi.org/10.5194/nhess-10-1697-2010
Merz, Bruno, Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J. C. J. H., Bates, P., et al. (2021). Causes, impacts and patterns of disastrous river floods. *Nature Reviews Earth & Environment*, *2*(9), 592–609. https://doi.org/10.1038/s43017-021-00195-3
Mitheu, F., Tarnavsky, E., Ficchì, A., Stephens, E., Cornforth, R., & Petty, C. (2023). The utility of impact data in flood forecast verification for anticipatory actions: Case studies from Uganda and Kenya. *Journal of Flood Risk Management*. https://doi.org/10.1111/jfr3.12911
Moftakhari, H. R., AghaKouchak, A., Sanders, B. F., Allaire, M., & Matthew, R. A. (2018). What Is Nuisance Flooding? Defining and Monitoring an Emerging Challenge. *Water Resources Research*, *54*(7), 4218–4227. https://doi.org/10.1029/2018WR022828
Mullen, L. (2016). Textreuse: Detect Text Reuse and Document Similarity. R package version 0.1.4.9000.
Nick, F. C., Sänger, N., van der Heijden, S., & Sandholz, S. (2023). Collaboration is key: Exploring the 2021 flood response for critical infrastructures in Germany. *International Journal of Disaster Risk Reduction*, *91*, 103710. https://doi.org/10.1016/j.ijdrr.2023.103710
Papagiannaki, K., Petrucci, O., Diakakis, M., Kotroni, V., Aceto, L., Bianchi, C., et al. (2022). Developing a large-scale dataset of flood fatalities for territories in the Euro-Mediterranean region, FFEM-DB. *Scientific Data*, *9*(1), 166. https://doi.org/10.1038/s41597-022-01273-x
Paprotny, D., Morales-Nápoles, O., & Jonkman, S. N. (2018). HANZE: a pan-European database of exposure to natural hazards and damaging historical floods since 1870. *Earth System Science Data*, *10*(1), 565–581. https://doi.org/10.5194/essd-10-565-2018
Qiang, Y., Huang, Q., & Xu, J. (2020). Observing community resilience from space: Using nighttime lights to model economic disturbance and recovery pattern in natural disaster. *Sustainable Cities and Society*, *57*, 102115. https://doi.org/10.1016/j.scs.2020.102115
Rahman, M. S., & Di, L. (2020). A Systematic Review on Case Studies of Remote-Sensing-Based Flood Crop Loss Assessment. *Agriculture*, *10*(4), 131. https://doi.org/10.3390/agriculture10040131
Reuters. (2021). Floods cut power to 200,000 households in western Germany. Retrieved from https://www.reuters.com/business/environment/floods-cut-power-200000-households-western-germany-2021-07-15/
Sanders, B. F., Schubert, J. E., Kahl, D. T., Mach, K. J., Brady, D., AghaKouchak, A., et al. (2022). Large and inequitable flood risks in Los Angeles, California. *Nature Sustainability*. https://doi.org/10.1038/s41893-022-00977-7
Schäfer, A., Mühr, B., Daniell, J., Ehret, U., Ehmele, F., Brand, J., et al. (2021). *Hochwasser Mitteleuropa, Juli 2021 (Deutschland)* (Vol. 2021). https://doi.org/10.5445/IR/1000135730.CEDIM
Schauff, D., & Schneider, D. (2021). Hochwasser drückt in die Straßen der Region – Verkehrsbehinderungen und Wasserschäden. *Rhein Zeitung*. Retrieved from https://www.rhein-zeitung.de/region/aus-den-lokalredaktionen/koblenz-und-region_artikel,-hochwasser-drueckt-in-die-strassen-der-region-verkehrsbehinderungen-und-wasserschaeden-_arid,2283957.html
Schreier, M. (2012). *Qualitative content analysis in practice*. Los Angeles: SAGE.
Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., et al. (2021). Weather and Climate Extreme Events in a Changing Climate. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. https://doi.org/10.1017/9781009157896.013
Serpico, S. B., Dellepiane, S., Boni, G., Moser, G., Angiati, E., & Rudari, R. (2012). Information Extraction From Remote Sensing Images for Flood Monitoring and Damage Evaluation. *Proceedings of the IEEE*, *100*(10), 2946–2970. https://doi.org/10.1109/JPROC.2012.2198030
Sieg, T., & Thieken, A. H. (2022). Improving flood impact estimations. *Environmental Research Letters*, *17*(6), 064007. https://doi.org/10.1088/1748-9326/ac6d6c
Sodoge, J., Madruga de Brito, M., & Kuhlicke, C. (2022). Automatized Drought Impact Detection Using Natural Language Processing. *WASSERWIRTSCHAFT*, *112*(S1), 30–31. https://doi.org/10.1007/s35147-022-1008-7
Sodoge, J., Kuhlicke, C., & de Brito, M. M. (2023). Automatized spatio-temporal detection of drought impacts from newspaper articles using natural language processing and machine learning. *Weather and Climate Extremes*,

100574. https://doi.org/10.1016/j.wace.2023.100574
Szyma, P., & Kajdanowicz, T. (2016). scikit-multilearn: A scikit-based Python environment for performing multi-label classification, *1*, 1–15.
Szymczak, S., Backendorf, F., Bott, F., Fricke, K., Junghänel, T., & Walawender, E. (2022). Impacts of Heavy and Persistent Precipitation on Railroad Infrastructure in July 2021: A Case Study from the Ahr Valley, Rhineland-Palatinate, Germany. *Atmosphere*, *13*(7), 1118. https://doi.org/10.3390/atmos13071118
Tapia-Silva, F.-O., Itzerott, S., Foerster, S., Kuhlmann, B., & Kreibich, H. (2011). Estimation of flood losses to agricultural crops using remote sensing. *Physics and Chemistry of the Earth, Parts A/B/C*, *36*(7–8), 253–265. https://doi.org/10.1016/j.pce.2011.03.005
The Guardian. (2022). After the floods: Germany's Ahr valley then and now – in pictures. Retrieved from https://www.theguardian.com/world/2022/jul/13/floods-then-and-now-photographs-germany-ahr-valley-flooding-disaster-july-2021#:~:text=None of the 18 bridges,on many of the buildings.
Thieken, A. H., Bubeck, P., Zenker, M. L., & Wutzler, B. (2022). *Strukturierte Auswertung der Dokumentationen zu allen Hochwassertodesopfern in Nordrhein-Westfalen im Juli 2021 und Herausarbeitung von Verbesserungspotenzialen in der Risikokommunikation und in den Warnprozessen anhand der Todesumstände und -ursachen sow*. Retrieved from https://www.landtag.nrw.de/home/der-landtag/ausschusse-und-gremien-1/untersuchungsausschusse/untersuchungsausschuss-ii-hochwa.html
UNDRR. (2021). *Regional Assessment Report on Disaster Risk in Latin America and the Caribbean*.
UNISDR. (2015). *Making development sustainable: the future of disaster risk management. Global assessment report on disaster risk reduction*. Geneva: UNISDR.
Vaughan, A. (2022). The day the floods came. *New Scientist*, *253*(3375), 8–9. https://doi.org/10.1016/S0262-4079(22)00308-6
Wikipedia. (2022). Hochwasser in West- und Mitteleuropa 2021. Retrieved from https://de.wikipedia.org/wiki/Hochwasser_in_West-_und_Mitteleuropa_2021
Yoon, J., Klassert, C., Selby, P., Lachaut, T., Knox, S., Avisse, N., et al. (2021). A coupled human–natural system analysis of freshwater security under climate and population change. *Proceedings of the National Academy of Sciences*, *118*(14). https://doi.org/10.1073/pnas.2020431118
Zeit Online. (2021). Rund 40.000 Menschen im Ahrtal von Hochwasser-Folgen betroffen. Retrieved from https://www.zeit.de/gesellschaft/2021-07/hochwasser-rheinland-pfalz-ahrweiler-folgen-ueberschwemmung

Supplementary Table S1. Newspaper outlets included in our search

Newspaper	Scale
Aachener Nachrichten	local
Aachener Zeitung	local
Aar-Bote	local
Abendzeitung	local
Achimer Kreisblatt	local
Alb Bote	regional
Aller-Zeitung	local
Allgemeine Zeitung Mainz-Rheinhessen	regional
Alsfelder Allgemeine	local
Badische Zeitung	regional
Barmstedter Zeitung	local
Bayerische GemeindeZeitung	regional
Bayerische Rundschau	regional
Bergedorfer Zeitung	local
Bergische Morgenpost	regional
Berliner Kurier	local
Berliner Morgenpost	local
Berliner Zeitung	local
Bersenbrücker Kreisblatt	local
Bonner General-Anzeiger	local
Börsen-Zeitung	national
Bote vom Haßgau	local
Bramscher Nachrichten	local
Braunauer Warte am Inn	local
Braunschweiger Zeitung	local
Bürstädter Zeitung	local
BZ	local
Chemnitzer Morgenpost	local
Christ und Welt	national
Coburger Tageblatt	local
Darmstädter Echo	local
Delmenhorster Kreisblatt	local
Der Insel-Bote	regional
Der Prignitzer	local
Der Rheintaler	regional
DER SPIEGEL	national
DER SPIEGEL online	national
Die Glocke	local
DIE KITZINGER	local
DIE WELT	national
Die Wirtschaft	national

DIE ZEIT	national
DIE ZEIT	national
DIE ZEIT online	national
Diepholzer Kreisblatt	local
Döbelner Allgemeine Zeitung	local
Dorstener Zeitung	local
Dresden am Wochenende	local
Dresdner Morgenpost	local
Dresdner Neueste Nachrichten	local
Eckernförder Zeitung	local
Eichsfelder Tageblatt	local
Elmshorner Nachrichten	local
Ems-Zeitung	regional
EXPRESS	national
FAZ Einspruch	national
FAZ Wirtschaftswissenschaft	national
Flensburger Tageblatt	local
FOCUS	national
FOCUS-MONEY	national
Frankenberger Allgemeine	local
Frankfurter Neue Presse	national
Frankfurter Rundschau	local
Fränkischer Tag	regional
Freie Presse	national
Freiheit	national
Fritzlar-Homberger Allgemeine	local
Geislinger Zeitung	local
Gelnhäuser Tageblatt	local
Gießener Allgemeine	local
Gießener Anzeiger	local
Gifhorner Rundschau	local
Glückstädter Fortuna	local
Göttinger Tageblatt	local
Haller Tagblatt	local
Halterner Zeitung	local
Hamburger Abendblatt	local
Hamburger Morgenpost	local
Handelsblatt	national
Handelsblatt Magazin	national
Handelsblatt Morning Briefing	national
Handelsblatt News am Abend	national
Handelsblatt online	national
Hannoversche Allgemeine Zeitung	local

HarzKurier	regional
Heidenheimer Zeitung	local
Heilbronner Stimme	local
Hennigsdorfer Generalanzeiger	local
Hessische Niedersächsische Allge…	regional
Hildesheimer Allgemeine Zeitung	local
Hochheimer Zeitung	local
Höchster Kreisblatt	local
Hofgeismarer Allgemeine	local
Hofheimer Zeitung	local
Hohenloher Tagblatt	local
Hohenzollerische Zeitung	local
Holsteinischer Courier	local
Husumer Nachrichten	local
Idsteiner Zeitung	local
Iserlohner Kreisanzeiger	local
Jüdische Allgemeine	national
Kieler Nachrichten	local
Kirner Zeitung	local
Kölner Stadt-Anzeiger	local
Kölnische Rundschau	local
Kreis-Anzeiger	local
Kreiszeitung	local
kulturSPIEGEL	national
Lampertheimer Zeitung	local
Landshuter Zeitung	local
Lausitzer Rundschau	regional
Lauterbacher Anzeiger	local
Le Monde diplomatique	national
Leipziger Volkszeitung	local
Lingener Tagespost	local
LiteraturSPIEGEL	national
Lübecker Nachrichten	local
Magdeburger General-Anzeiger	local
Magdeburger Volksstimme	local
Main-Post	regional
Main-Spitze	regional
Main-Taunus-Kurier	regional
Märkische Allgemeine	regional
Märkische Oderzeitung	regional
Märkische Zeitung - Gransee-Zeitung	regional
Märkische Zeitung - Ruppiner Anz…	regional
Meller Kreisblatt	local

Melsunger Allgemeine	local
Meppener Tagespost	local
Metzinger Uracher Volksblatt	local
Mittelbayerische Zeitung	regional
Mitteldeutsche Zeitung	regional
mittwochSZ und freitagSZ	national
Münchner Merkur	local
Mündener Allgemeine	local
Münsterland Zeitung	regional
Nahe-Zeitung	regional
Nassauische Neue Presse	local
Neckar-Chronik	local
Neue Osnabrücker Zeitung	local
Neue Presse	national
Neue Ruhr/Neue Rhein Zeitung	regional
Neue Westfälische	regional
Neue Württembergische Zeitung	regional
Neuss-Grevenbroicher Zeitung	regional
Nordbayerischer Kurier	regional
Norddeutsche Neueste Nachrichten	regional
Norddeutsche Rundschau	local
Nordfriesland Tageblatt	regional
Nordkurier	regional
Nordwest-Zeitung	regional
Northeimer Neueste Nachrichten	local
Nürnberger Nachrichten	local
Nürnberger Zeitung	local
Oberhessische Presse	regional
Oberhessische Zeitung	regional
Öffentlicher Anzeiger	local
Oranienburger Generalanzeiger	local
Oschatzer Allgemeine Zeitung	local
Osterländer Volkszeitung	local
Ostholsteiner Anzeiger	local
Ostsee-Zeitung	regional
Ostthüringer Zeitung	regional
Passauer Neue Presse	regional
Peiner Allgemeine Zeitung	regional
Pfälzischer Merkur	regional
Pinneberger Tageblatt	local
Quickborner Tageblatt	local
Regensburger Zeitung	local
Reutlinger General-Anzeiger	local

Reutlinger Nachrichten	local
Rhein-Hunsrück-Zeitung	regional
Rheinische Post	regional
Rhein-Lahn-Zeitung	regional
Rhein-Zeitung	regional
Rieder Volkszeitung	local
Rotenburg-Bebraer Allgemeine	local
Rotenburger Kreiszeitung	local
Ruhr Nachrichten	regional
Rundschau für den schwäbischen Wald	regional
Rüsselsheimer Echo	local
Saale Zeitung	local
Saarbrücker Zeitung	local
Sächsische Zeitung	regional
Salzgitter-Zeitung	local
Schaffhauser Nachrichten	local
Schenefelder Tageblatt	local
Schlei-Bote	local
Schleswiger Nachrichten	regional
Schleswig-Holsteinische Landesze…	regional
Schwäbische Zeitung	regional
Schwäbisches Tagblatt	regional
Schwälmer Allgemeine	local
Schweinfurter Tagblatt	local
Schweriner Volkszeitung	local
Segeberger Zeitung	local
Solinger Morgenpost	local
Sollinger Allgemeine	local
SPIEGEL Bestseller	national
SPIEGEL Plus	national
SPIEGEL special	national
SPIEGEL Start	national
Stern	national
Stormarner Tageblatt	local
Straubinger Tagblatt	local
Südkurier	regional
SÜDWEST PRESSE	regional
Sulinger Kreiszeitung	local
Sylter Rundschau	local
Taunus Zeitung	regional
taz die tageszeitung	national
Thedinghäuser Zeitung	local
Thüringer Allgemeine	regional

Thüringische Landeszeitung	regional
Torgauer Zeitung	local
Trierischer Volksfreund	local
Uetersener Nachrichten	local
uniSPIEGEL	national
Usinger Anzeiger	local
Verdener Aller-Zeitung	local
Volksblatt Würzburg	local
Volkszeitung Schweinfurt	local
Wedel-Schulauer Tageblatt	local
WELT am SONNTAG	national
WELT KOMPAKT	national
WELT ONLINE	national
Westdeutsche Allgemeine Zeitung	regional
Westdeutsche Zeitung	regional
Westerwälder Zeitung	regional
Westfalen-Blatt	regional
Westfalenpost	regional
Westfälische Rundschau	regional
Wetterauer Zeitung	local
Wiesbadener Kurier	local
Wiesbadener Tagblatt	local
Wildeshauser Zeitung	local
Wilstersche Zeitung	local
Wirtschaft in Sachsen	regional
Wirtschaftszeitung	national
Wittlager Kreisblatt	local
Witzenhäuser Allgemeine	local
Wolfenbütteler Zeitung	local
Wolfhager Allgemeine	local
Wolfsburger Allgemeine Zeitung	local
Wolfsburger Nachrichten	local
Wormser Zeitung	local
ZEIT Campus	national
ZEIT Geschichte	national
ZEIT Studienführer	national
ZEIT Wissen	national

Supplementary Table S2. Topic model results sorted according to the topic prevalence

Topic	Prevalence	Top terms
1	4.88	spend, euro, betroff, geld, unterstutz, aktion, helf, verein, hilf, opf, spendenkonto, spendenaktion, hilft, gespendet, flutopf, flutkatastroph, iban, gesammelt, summ, unterstutzt, club, betrag, vorsitz, finanziell, hochwasserkatastroph
2	4.03	sich, halt, situation, entscheid, lang, verantwort, nach, grund, leut, arbeit, passiert, probl, find, gesprach, erfahr, schwierig, problem, nehm, geld, polit, ding, fall, fehl, zeigt, burg
3	3.29	haus, famili, stand, kell, flut, wohnung, alt, strass, jahrig, nacht, paar, frau, met, gluck, leb, nachbarn, freund, rett, leut, strom, schlamm, schlimm, wohnt, zerstort, nach
4	3.21	einsatz, thw, helf, einsatzkraft, drk, kraft, technisch, ehrenamt, hilfswerk, technisch_hilfswerk, rot, kreuz, rot_kreuz, ortsverband, unterstutz, katastrophengebiet, fahrzeug, hochwassergebiet, aufgab, nurburgring, versorg, arbeit, maltes, dlrg, hilfsorganisation
5	2.99	hochwasserschutz, massnahm, flach, planung, geplant, bereich, plan, uberschwemmungsgebiet, bau, projekt, baugebiet, bach, gebiet, zustand, grundstuck, schutz, gewass, gutacht, umsetz, notwend, aktuell, neu, besteh, umgesetzt, bebau
6	2.98	starkreg, massnahm, starkregenereignis, ereignis, schutz, burg, kart, kommun, bereich, uberflut, gefahr, konzept, gefahrdet, schad, folg, vorsorg, betroff, extrem, erstell, treff, moglich, gewass, stark, uberschwemm, erstellt
7	2.7	kell, strass, feuerwehr, einsatz, uberflutet, starkreg, vollgelauf, betroff, unwett, voll, stark, stand, polizei, nacht, vollgelauf_kell, lief, wassermass, reg, gesperrt, baum, regenfall, heftig, verletzt, zahlreich, uberschwemm
8	2.56	gemeind, burgermeist, gemeinderat, massnahm, burg, sitzung, kost, euro, jung, konzept, hochwasserschutz, burgermeisterin, prozent, ortsteil, rat, forder, berichtet, einstimm, beschloss, rathaus, gesprach, verwalt, stellt, offent, bauhof
9	2.39	strass, starkreg, kanal, kell, haus, regenwass, kanalisation, grab, anwohn, grundstuck, reg, schutz, beck, flach, stark, probl, rohr, uberschwemm, regenruckhaltebeck, uberflut, bau, aufnehm, starkregenereignis, regelmass, wassermass
10	2.36	met, arbeit, hochwasserschutz, hoh, lang, bereich, deich, bau, euro, projekt, neu, geplant, gebaut, alt, strass, schutz, plan, kost, million, nach, planung, damm, massnahm, million_euro, fertig
11	2.32	spend, helf, sachspend, transport, benotigt, aktion, lkw, betroff, hilfsgut, kleidung, verfug, hilf, voll, dringend, gebraucht, fahr, facebook, kontakt, lebensmittel, gebracht, hilfsbereitschaft, organisiert, sach, bring, gesammelt
12	2.3	rheinland, pfalz, rheinland_pfalz, westfal, nordrhein, nordrhein_westfal, westfal_rheinland, betroff, pfalz_nordrhein, hochwasserkatastroph, flutkatastroph, nrw, region, katastroph, gebiet, bundesland, verheer, nrw_rheinland, west, unwett, leb, teil, angab, uberschwemm, flut
13	2.3	helf, hilf, freiwill, ahrtal, einsatz, arbeit, leut, dankbar, schlamm, haus, hand, team, freiwill_helf, fahr, unterstutz, grupp, ehrenamt, geholf, gebraucht, ahrweil, organisiert, ess, jahrig, schwer, katastrophengebiet
14	2.25	spd, cdu, antrag, grun, fraktion, verwalt, sitzung, ausschuss, hochwasserschutz, burg, stadtrat, fdp, fraktionsvorsitz, polit, rat, fordert, berat, beschloss, nach, anfrag, antwort, punkt, tagesordn, gestellt, stadtverwalt
15	2.24	veranstalt, musik, konzert, besuch, gast, abend, band, benefizkonzert, buhn, programm, erlos, publikum, corona, opf, zugun, spend, ide, musikal, zweck, kuch, getrank, teilnehm, flutopf, kunstl, auftritt
16	2.18	schad, gebaud, betroff, kell, komplett, stand, beschadigt, gezog, anlag, bod, getroff, stark, zerstort, mitleidenschaft, raum, schwer, uberflutet, mitleidenschaft_gezog, betrieb, arbeit, neu, angerichtet, gluck, entstand, dau
17	2.15	bach, bruck, met, ufer, flut, wassermass, strass, baum, fluss, schad, reissend, tief, hoh, richtung, dorf, tal, schlimm, nah, stand, stark, gefahr, hang, breit, schwer, uberflutet
18	2.15	betroff, katastroph, helf, flut, verlor, hilf, haus, leb, zerstort, ausmass, erlebt, zerstor, schlimm, leid, situation, schwer, opf, hilfsbereitschaft, unterstutz, solidaritat, angehor, gebiet, region, existenz, verheer
19	2.01	lit, pro, quadratmet, pro_quadratmet, starkreg, dwd, lit_pro, reg, stund, wetterdien, gewitt,

		niederschlag, unwett, teil, grad, wetterdien_dwd, erneut, meteorolog, regenmeng, heftig, stark, lit_reg, lokal, schau, nacht
20	1.97	betroff, mail, per, hilf, burg, onlin, per_mail, antrag, information, nrw, telefon, unterstutz, angebot, verfug, person, burgerinn, anmeld, soforthilf, burgerinn_burg, meld, erhalt, berat, land, eingerichtet, moglich
21	1.91	euro, million, million_euro, schad, land, soforthilf, hoh, kost, geld, betroff, haushalt, bund, prozent, milliard, milliard_euro, ausgezahlt, kommun, antrag, infrastruktur, summ, verfug, finanziell, entstand, gestellt, gebaud
22	1.73	feuerwehr, einsatz, freiwill, feuerwehrleut, freiwill_feuerwehr, einsatzkraft, kamerad, fahrzeug, wehr, pump, stellvertret, feuerwehrmann, kraft, hilf, stund, kreisbrandmeist, kommandant, unterstutz, wehrfuhr, landkreis, einsatzleit, sandsack, loschzug, minut, kell
23	1.59	spiel, verein, platz, fussball, sv, mannschaft, fc, train, minut, team, tus, zuschau, sport, tor, sportplatz, sc, tsv, saison, start, parti, sg, turni, training, sportverein, kilomet
24	1.5	met, pegel, wasserstand, pegelstand, rhein, lag, erreicht, zentimet, hoh, fluss, stark, steigend, stand, regenfall, gefahr, stieg, erwartet, uberschritt, situation, nacht, hochststand, anstieg, prognos, ufer, meldestuf
25	1.47	strass, bruck, bahn, gesperrt, streck, schad, verkehr, richtung, beschadigt, fahr, autobahn, dau, zug, gleis, sperrung, lang, bahnhof, zerstort, arbeit, kilomet, bereich, stark, koln, befahrbar, fahrbahn
26	1.44	grun, polit, klimawandel, klimaschutz, kommun, folg, bund, land, fordert, stadt, partei, ziel, zukunft, co, flach, stark, klima, nachhalt, forder, konkret, kunftig, anpass, klimakris, prozent, energi
27	1.40	schad, versicher, versich, versichert, haus, gebaud, hoh, gdv, elementarschad, prozent, milliard, starkreg, uberschwemm, elementarschadenversicher, verbraucherzentral, euro, milliard_euro, versicherungswirtschaft, miet, hausbesitz, hausrat, gesamtverband, sturm, gesamtverband_versicherungswirtschaft, pflichtversicher
28	1.39	fluss, gewass, hochwasserschutz, bach, flach, natur, schutz, umwelt, gefahr, gebiet, ereignis, extrem, gross, uberschwemm, betroff, naturschutz, renaturier, tal, ahr, raum, landesamt, katastroph, ufer, land, klein
29	1.37	laschet, cdu, ministerprasident, merkel, armin, spd, armin_laschet, scholz, besuch, drey, angela, kanzlerkandidat, angela_merkel, hilf, betroff, nrw, olaf, laschet_cdu, malu_drey, malu, polit, ministerprasident_armin, steinmei, ministerprasidentin, baerbock
30	1.35	corona, pandemi, verein, vorsitz, mitglied, veranstalt, corona_pandemi, termin, geimpft, vorstand, stattfind, zahl, spend, neu, blutspend, drk, person, teilnehm, test, alt, aktiv, aktuell, arbeit, impfung, genes
31	1.32	unternehm, mitarbeit, betrieb, betroff, kund, geschaftsfuhr, kolleg, arbeit, firma, standort, region, werk, ihk, gmbh, beschaftigt, firm, produktion, unterstutz, filial, handwerk, mitarbeiterinn, maschin, stadtwerk, prozent, stark
32	1.29	haus, bewohn, strom, lag, nacht, evakuiert, feuerwehr, rett, krankenhaus, stund, strass, einsatz, wassermass, helf, evakuier, bundeswehr, sich, hubschraub, verlass, boot, gerettet, gebaud, polizei, person, gebracht
33	1.26	sir, warnung, bevolker, warn, katastrophenschutz, app, feuerwehr, nina, gewarnt, katastroph, landkreis, kommun, apps, katastrophenfall, gefahr, digital, burg, alarmier, information, leitstell, warn_app, informiert, ernstfall, warn_apps, land
34	1.26	ahrweil, bad, neuenahr, bad_neuenahr, ahrtal, ahr, neuenahr_ahrweil, rheinland, zerstort, pfalz, rheinland_pfalz, sinzig, flut, altenahr, dernau, betroff, landkreis_ahrweil, flutkatastroph, wiederaufbau, koblenz, stark, landkreis, pfalzisch, bruck, rheinland_pfalzisch
35	1.21	bod, landwirt, wald, landwirtschaft, baum, feld, flach, reg, stark, pflanz, tier, natur, ernt, trock, bau, hektar, wies, gart, nass, forst, bib, mais, somm, halt, prozent
36	1.13	bund, land, wiederaufbau, milliard, milliard_euro, bund_land, euro, unternehm, betroff, schad, hilf, infrastruktur, bundestag, geld, zerstort, fond, flutkatastroph, berlin, spd, bundesregier, beschloss, finanziell, firm, gesetz, wiederaufbaufond
37	1.10	berchtesgad, uberschwemm, unwett, land, betroff, regenfall, haus, lag, heftig, berchtesgad_land, teil, mindest, landkreis, passau, stark, schwer, osterreich, strass, angab, leb, ums, region, erdrutsch, ums_leb, tot

38	1.06	kind, famili, jugend, elt, jung, angebot, kind_jugend, grupp, ide, web, unterstutz, betreu, erwachs, team, aktion, alt, schon, gemeinsam, madch, organisiert, zimmermann, besuch, urlaub, kontakt, schul
39	1.05	mull, sperrmull, tonn, trinkwass, entsorg, abfall, ol, betroff, meng, schlamm, contain, anlag, entsorgt, gefahr, klaranlag, heizol, stoff, zusatz, mitarbeit, strass, schadstoff, deponi, wasserwerk, beseit, sprecherin
40	1.02	kirch, st, evangel, kirchengemeind, pfarr, gemeind, kathol, opf, flutkatastroph, gottesdien, andacht, glock, caritas, betroff, evangel_kirch, diakoni, laut, gebet, zeich, opf_flutkatastroph, schutz, kirchenkreis, evangel_kirchengemeind, bistum, gedenk
41	1.01	landrat, pfohl, staatsanwaltschaft, cdu, koblenz, ahrweil, flutkatastroph, fahrlass, ahrtal, rheinland, lewentz, mainz, krisenstab, jurg, verantwort, ermittl, jurg_pfohl, anfangsverdacht, untersuchungsausschuss, warnung, landtag, pfalzisch, rheinland_pfalzisch, innenminist, kreisverwalt
42	1.01	sachs, brandenburg, elb, landkreis, elst, anhalt, land, thuring, dresd, mz, sachs_anhalt, niedersachs, sachsisch, saal, west, lausitz, schwarz, grimma, holstein, region, dobeln, wittenberg, schleswig, cdu, gera
43	0.99	landkreis, bay, wasserwirtschaftsamt, donau, bayer, pold, landratsamt, csu, munch, freistaat, regensburg, betroff, deggendorf, hochstadt, aisch, glaub, cham, bamberg, hassberg, neustadt, wurzburg, coburg, nurnberg, kelheim, gemeind
44	0.97	nrw, dusseldorf, cdu, reul, landesregier, ess, hein, krisenstab, westfal, nordrhein, nordrhein_westfal, hein_ess, landtag, kommun, innenminist, herbert, flutkatastroph, land, herbert_reul, innenministerium, ministerin, innenminist_herbert, spd, ursula, scharrenbach
45	0.97	talsperr, kubikmet, sekund, pro, pro_sekund, huckeswag, damm, beck, wupp, bad, wupperverband, kubikmet_pro, wehr, oeynhaus, bad_oeynhaus, million, million_kubikmet, herford, ruckhaltebeck, werr, wipperfurth, uberlauf, hochwasserschutz, see, stause
46	0.97	prozent, klimawandel, region, somm, extrem, haufig, grad, stark, starkreg, folg, berlin, hoh, niederschlag, warm, hannov, wetterlag, zunehm, trock, zahl, temperatur, tief, lag, luft, west, durr
47	0.95	tier, frau, polizei, jahrig, schaf, jung, mann, hund, pferd, tierheim, simon, alt, flut, gefund, leb, besitz, fisch, katz, auto, gerettet, vermisst, tot, sich, fuss, sozial
48	0.94	stolberg, eschweil, aach, gast, hotel, campingplatz, innenstadt, platz, corona, urlaub, stadteregion, restaurant, vicht, geschaft, roetg, tourist, mitarbeit, haas, ind, betrieb, eschweil_stolberg, geoffnet, mulartshutt, tourismus, betreib
49	0.88	pet, han, burgermeist, heinz, jurg, schmidt, karl, vorsitz, mey, mend, ortsvorsteh, michael, berichtet, markt, wolf, mitglied, alt, stellvertret, diet, karl_heinz, thomas, meier, dr, balv, wolfgang
50	0.86	vg, verbandsgemeind, prum, bitburg, land, may, ortsburgermeist, koblenz, wittlich, gerolstein, rhein, ahr, ortsgemeind, kyll, gemeind, burg, mosel, eifelkreis, vulkaneifel, landkreis, ort, kreisverwalt, sgd, neuwied, bitburg_prum
51	0.85	sandsack, feuerwehr, einsatz, arnsberg, iserlohn, altena, sund, burg, sack, sand, hem, markisch, lag, herford, bauhof, olp, frondenberg, gefullt, mitarbeit, mind, einsatzkraft, lubbeck, bernd, betroff, hoxt
52	0.81	schul, kita, kind, grundschul, gymnasium, gebaud, raum, elt, kindergart, einricht, kitas, klass, lehr, contain, sporthall, kindertagesstatt, genutzt, schulerinn, realschul, standort, rund, lauf, steff, schulerinn_schul, turnhall
53	0.79	seehof, bund, land, katastrophenschutz, bevolkerungsschutz, bundesamt, bbk, katastrophenhilf, warnung, bundesamt_bevolkerungsschutz, bevolkerungsschutz_katastrophenhilf, horst, cell, horst_seehof, csu, schust, bundesinnenminist, behord, bevolker, syst, bundesinnenminist_horst, verantwort, seehof_csu, bund_land, polit
54	0.79	hag, gevelsberg, ruhr, herdeck, soling, bochum, wuppertal, witt, remscheid, wupp, wett, hohenlimburg, ennep, unterburg, leverkus, schad, ennepetal, schwelm, ennep_ruhr, burg, dahlhaus, dortmund, bergisch, flut, betrieb
55	0.79	bad, landkreis, rheinland, pfalz, rheinland_pfalz, wurttemberg, bad_wurttemberg, hess, land, frankfurt, wiesbad, nordrhein, westfal, nordrhein_westfal, markisch, betroff, regierungsprasidium, hessisch, ulm, braunsbach, region, reutling, schwabisch, ahrweil, hall

56	0.74	hamburg, museum, histor, kunst, alt, buch, geschicht, akt, werk, ausstell, haus, archiv, kais, dokument, sammlung, arbeit, erhalt, objekt, fotos, les, depot, besuch, koln, leb, rett
57	0.74	vermisst, ort, haus, rheinland, polizei, tot, westfal, nordrhein_westfal, nordrhein, betroff, rheinland_pfalz, pfalz, mindest, lag, flut, spd, hag, starb, schuld, einsatzkraft, drey, teilt, altena, dutzend, unwett
58	0.70	ess, ruhr, duisburg, deich, mulheim, wesel, rhein, niederrhein, klev, xant, moer, nier, schiff, dinslak, lun, nrw, bezirksregier, deichverband, willich, lipp, viers, rheinberg, niederland, oberhaus, rees
59	0.67	bonn, sieg, rhein, swisttal, rheinbach, rhein_sieg, koln, heimerzheim, meckenheim, odendorf, betroff, bornheim, gemeind, euskirch, ga, swist, lohmar, alft, wachtberg, mert, flutkatastroph, kreis, flut, anzeig, burg
60	0.66	erftstadt, erft, bless, rur, dur, rhein, rhein_erft, koln, wassenberg, heinsberg, wver, wasserverband, burg, julich, kreis, ophov, ind, kiesgrub, erftstadt_bless, haus, geilenkirch, wurm, wasserverband_rur, zweibruck, grub
61	0.65	euskirch, bad, munstereifel, bad_munstereifel, schleid, langenberg, erft, gutersloh, bielefeld, kall, paderborn, flut, erftverband, gemund, neuss, mechernich, velbert, betroff, zerstort, weilerswist, flutkatastroph, ram, schloss, hellenthal, nrw
62	0.62	trier, winz, ahr, ehrang, wein, kordel, saarburg, weingut, flut, flasch, mayschoss, ahrtal, trier_saarburg, weinberg, region, sau, zerstort, menk, betroff, trier_ehrang, traub, mosel, konz, balt, kyll
63	0.57	leichling, bergisch, hild, erkrath, mettmann, rheinisch, schulz, wagn, rating, rosrath, thiel, langenfeld, itt, brw, haan, wahl, dusseldorf, leverkus, rhein, gladbach, gruit, berg, wermelskirch, verein, rheinisch_bergisch
64	0.53	haus, schuld, ahr, ahrtal, altenburg, beschadigt, zerstort, altenahr, frau, gasp, abgeriss, bernd, schwer, flut, strass, alt, dorf, wassermass, katastroph, schwer_beschadigt, helf, schmitz, einwohn, insul, achim
65	0.48	dusseldorf, erding, dussel, landkreis, st, vorpomm, dorf, saarland, siedlung, betroff, mecklenburg, jost, mv, saarbruck, gerresheim, ostparksiedl, rostock, grimm, freising, anwohn, mecklenburg_vorpomm, ingbert, rug, oberdorf, st_ingbert

Supplementary Table S3. Topology of flood impact classes and number of flood impact statements (FIS)

Class	Sub-class	Description	N of FIS
Individuals	1.1 Physical injuries	Includes both severe and mild physical injuries resulting directly from floods or secondary events (e.g. car crashes)	620
	1.2 Loss of life	Includes the loss of life caused by floods, including drowning and accidents	1933
	1.3 Mental health	Includes articles about persons suffering from trauma due to flood events or articles that describe mental health services available for flood victims	345
	1.4 Temporary displacement and homelessness	Articles in this class describe situations in which persons lost their homes and are forced to live temporarily elsewhere (e.g. friends, family, shelters) or that became homeless	594
	1.5 Loss of livelihood or income	Describe situations in which persons lost their income due to unemployment or short-time work or that lost their livelihood in a way that threatens their existence	289
Critical infrastructure	2.1 Electricity, gas and heating	Articles in this class describe situations in which electricity, gas or heating distribution were disrupted	1580
	2.2 Transport and mobility	Articles in this class describe situations in which transport and mobility were disrupted, including bridges collapses	3038
	2.3 Water supply	Articles in this class describe situations in which water supply was disrupted	501
	2.4 Water contamination	Includes articles about water contamination due to bacteria, toxins, pesticides and oil. It also encompasses articles that request communities to boil the water due to possible water contamination	214
	2.5 Sewage systems	Articles in this class describe situations in which sewage systems were overflown or sewage treatment plants were impaired	227
	2.6 Waste and debris management	Includes situations in which waste and debris caused disruption, and their disposal and management were impaired	806
	2.7 IT, communication and digital infrastructure	Articles in this class describe situations in which IT, communication and digital infrastructure were disrupted	233
	2.8 Education	Includes articles about schools that were physically damaged. It also describes situations in which students could not have classes due to the floods	335
	2.9 Child care	Includes articles about kindergartens that were physically damaged or that had their operation compromised	186
Economic	3.1 Need for economic support or monetary donations	Describe the need for economic support via governmental aid programs (e.g. disaster relief). It also includes monetary donations made to affected persons and companies	2746
Physical damage	4.1 Damage to buildings	Includes direct damages to buildings and constructions	3777
	4.2 Damage to private assets	Includes direct damages to private assets (e.g. furniture, cars)	1210
Economic sectors	5.1 Agriculture and livestock	Articles in this class describe impacts on agriculture (e.g. flooded cultivated areas, crop failures, loss of machinery) and livestock production (e.g. lack of feed for animals, dead livestock)	338
	5.2 Industry and commerce	Includes damages to industries and commerce	917
	5.3 Tourism and gastronomy	Includes damages to tourism and gastronomy (e.g. hotels and restaurants closure)	163

Supplementary Table S3. Dictionary used to classify the newspaper articles

<table>
<tr><th>Class</th><th>Sub-class</th><th>Keywords (| = OR, & = AND, .* = wildcard)</th></tr>
<tr><td rowspan="5">1. Individuals</td><td>1.1 Physical injuries</td><td>leichtverletzt.* | körperverletzung | schnittverletzung.* | personenschäden | unterkühlung | verletz.* & not (keine.verletzten | nicht.verletzt | unverletzt | ((verletz.* | personenschäden) & (niemand | nicht | tier.*)) | verletzungsgefahr)</td></tr>
<tr><td>1.2 Loss of lives</td><td>leichenbergung | opferzahlen | todesopf.* | tode | todesf.* | tot | tote | toten | tötung | ums.leben | ((bewohner.* | feuerwehrm* | frau | jähriger | mann | mensch.* | mutter | person.* | vater) & (ertrunken | ertrinken | gestorben | getötet | leben.gekostet | starb.* | sterben | stirb.*)) | ((angehörige | bekannte | familienmitglieder | freunde | leben | menschenleben) & (verlor.* | verlieren)) & not (keine.toten | keine.tode.* | ((ertrunken | ertrinken | gestorben | getötet | opferzahlen | starb.* | sterben | stirb.* | todesfälle | todesfall.* | todesopf.* | todesf.* | tot | tote | toten) & (coronavirus | fische | .*tier.*)) | (ums.leben & niemand))</td></tr>
<tr><td>1.3 Mental health</td><td>akuttraumat.* | panikattacken | posttraumatisch.* | traumatis.* | ((emotional | psycholog.* | psychotherapeutische | psychosoziale | psychisch.* | seelische) & (aufgewühlt | behandlung | belastung.* | belastet.* | betreuung | beratung.* | effekt.* | hilfe.* | nachsorge | notfallversorgung | problem.* | soforthilf.* | trauma.* | unterstütz))</td></tr>
<tr><td>1.4 Temporary displacement and homelessness</td><td>notuntergebracht | notunterkunft | (muss & ausziehen) | ((auto.* | feldbett.* | ferienwohnung.* | kirche | turnhalle.*| verwandten | zelt.*) & (.*schlaf.* | übernacht.*)) | ((obdachlos | heimatlos) & (derzeit | geworden.* | gemacht | werden | wurden)) | ((lebt.* | kostenlos.* | verfügung) & ferienwohnung.*) | ((dach | gebäude | haus | hause | hauses | häus.* | häuser | heim | wohnort | wohnhäuser | wohnung.*) & (unbewohnbar.* | über.dem.kopf | verlieren | verloren))</td></tr>
<tr><td>1.5 Loss of livelihood or income</td><td>arbeitsausfäll.* | einkommensverluste | existenzbedrohende.* | existenzgefährdung | existenzgrundlage | ((existenz.* | lebensgrundlage.* | zahlungsfähigkeit) & (an.den.rand | bedrohen | genommen | not | überschwemmt | verlier.* | verloren | vernichtet | zerstör.*)) | ((arbeitslos.* | kurzarbeit) & (angemeldet | erhöht | momentan | musste | stiegen))</td></tr>
<tr><td rowspan="3">2. Criticalinfrastructure</td><td>2.1 Energy, gas and heating</td><td>noch.strom | notstrom.* | stromausf.* | stromaggregat.* | stromlos | ((abbrachen | abgeschalt.* | abgeschnitt.* | abgestellt | abgesoffen | ausfäll.* | ausgefallen | beeinträchtigt | betrifft.* | betroffen | berflutete | beschädig.* | fall.* | fehlend.* | fehlt.* | fielen | gefallen | gestört.* | gesperrt | hauptproblem | hergestellt | herzustellen | kaputt | kein.*| ohne | problem | provisorische.* | niederspannungsversorgung | nicht.mehr | noch.nicht | notversorgung |.*schäde.* | sperren | störungen | überspüllt.* | unterspült.* | unterbroch.* | ungebrochen | weg | weggespült.* | wieder | wiederherstellung | zerbroch.* | zerlegt.* | zerstör.* | zusammengebrochen) & (biogas | energie | energieversorgung.* | erdgas.* | gas | gasleitungen | gasversorgung.* | .*heizung.* | .*strom.*))</td></tr>
<tr><td>2.2 Transport and mobility</td><td>.*behelfsbrücke.* | hilfsbrücke | notbrücke.* | straßensperrungen | schienenersatzverkehr | weggeschwemmte.wege | ((abbrachen | abgeschalt.* | abgesackt.* | abgeschnitt.* | ausfäll.* | bedeckt | beeinträchtigt | berflutete | beschädig. * | betrifft.* | betroffen | blockiert.* | demoliert | eingeschränkt | fall.* | fehlend.* | fortgespült | gefallen | gestört.* | gesperrt | getroffen | geflutet | nicht.befahrbar | rissen | .*schäde.* | sperren | .*sperrung.* | störung.* | überflutet,* | unbefahrbar | überschwemmt | überspüllt.* | unfäll.* | unfall | unpassierbar | unterspült.* | unterbroch.* | unter.wasser | vernichtet | verschoben | weg | weggerissen | weggespült.* | wiederaufgebaut | zerbroch.* | zerlegt.* | zerstör.* | zusammengebrochen) & (autobahn.* | bahn | bahnen | .*bahnstrecke.* | .*brück.* | busse.* | deutsche.bahn | eisenbahnschienen | fahrbahn.* | fußgang | fußweg | gleis.* | kleinbusse.* | öpnv | pendelstrecke | .*radwege | regionalverkehrsstrecken | schienen | schiff | .*straße.* | strecken | transport.* | .*verkehr.* | wegenetz | züg | züge | zügen)) | ((straße | wege) & blockiert)</td></tr>
<tr><td>2.3 Water supply</td><td>kein.wasser | kein.trinkwasser | kein.fließendes.wasser | kein.leitungswasser | ohne.wasser | trinkwassertank.* | wasserfässer | wasserwagen | weder.wasser | (((abbrachen | abgeschalt.* | abgeschnitt.* | ausfäll.* | ausgefallen | beeinträchtigt.* |</td></tr>
</table>

		berflutete | beschädig.* | betrifft.* | betroffen.* | beschädig. * | chlor.* | demoliert | eingeschränkt | fall.* | fehlend.* | gefallen | gesperrt | geflutet | gekappt | getroffen | gestört.* | gespült | ohne.* | mangelte | notversorgung | provisorisch.* | rohrbrüche | schäde.* | sperren | störung.* | tank.* | überspüllt.* | unterspült.* | unterbroch.* | vernichtet | versorgung | weggerissen | weggespült.* | wiederherstellung | zerbroch.* | zerlegt.* | zerstör.* | zusammengebrochen) | (funktioniert & nicht)) & (gewässerstruktur.* | leitungswasser | trinkwasser.* | wasserleitungen | .*wasserversorgung.* | ((trink | versorgung) & wasser)))
	2.4 Water contami-nation	((bakterien.* | fäkalien.* | giftsstoffen | heizöl | keime | kontaminiert.* | öl | parasiten | pilze | pflanzenschutzmittel | schadstoffe | verunreinigt) & (gewässer | leitungswasser | trinkwasser | wasser)) | (.*wasser.* & (abgekocht.* | abkochgebot.* | abkochen | abzukochen)) & not (nicht & öl & verunreinigt)
	2.5 Sewage systems	((abgesackt.* | abgeschnitt.* | ausfäll.* | beeinträchtigt | berflutete | beschädig.* | betrifft.* | betroffen.* | demoliert | erschöpft | fortgespült | geflutet | gesperrt | gestört.* | getroffen | notversorgung | schäde.* | sperren | störung.* | überschwemmt | überschwemmung | überlastet.* | überspüllt.* | unterbroch.* | übergelauf.* | überfüllt.* | vernichtet | weggerissen | weggespült.* | wiederaufbau | zerbroch.* | zerlegt.* | zerstör.* | zusammengebrochen | zusammengeschlossen) & (abwasser.* | kanalisation | kläranlag.* | schmutzwasser.*)) | keine.spülung
	2.6 Waste and debris management	abfallbeseitigung | entsorgungsproblem.* | geröll.* | hochwassersperrmüll | hochwasserabfällen | hochwassermüll | müllhaufen | müllberge | müllflut | .*müllmengen | sperrmüll-berge | sperrmüllabfuhr | (überall & müll) | ((abgesackt.* | abgeschnitt.* | aufräumen | ausfäll.* | beeinträchtigt | berflutete | entsorg.* | fall.* | fehlend.* | fielen | fortgespült | gestört.* | gesperrt | schäde.* | sonderabfuhr | sperren | störung.* | überspüllt.* | tonnen | unterbroch.* | übergelauf.* | überfüllt.* | vernichtet | versorgung |weggerissen | weggespült.* | zerbroch.* | zerlegt.* | zerstör.* | zusammengebrochen | zusammengeschlossen) & (abfall | müll | müllflächen | mülleimer | müllberge.* | müllentsorgung.* | müllsäcken | schlamm | schrott | schutt | sperrmüll.* | trennsystem | verbrennungsanlag.*))
	2.7 IT, commu-nication and digital infrastruc-ture	kein.mobilfunknetz | kein.handynetz | keinen.handyempfang | ((abbrachen | abgeschalt.* | abgesackt.* | abgeschnitt.* | ausfäll.* | ausgefallen | beeinträchtigt | berflutete | beschädig.* | betrifft.* | betroffen | demoliert | fall.* | fehlend.* | fielen | fortgespült | geflutet | gesperrt | gestört.* | getroffen | heruntergefahrene | (nicht & erreich.*) | notversorgung | ohne | .*schäde.* | sperren | störung.* | überschwemmt | überspüllt.* | unterbroch.* | vernichtet | verschoben | weggerissen | weggespült.* | zerbroch.* | zerlegt.* | zerstör.* | zusammengebrochen.*) & (digitale.infrastruktur | festnetz.* | handynetz.* | handyempfang | internetverbindungen | it | kommunikation.* | kommunikationsstruktur | kommunikationsanlagen | telefone | telefonanlage | telefonnetz.* | telekom | telekommunikation | mobilfunk.*)) | ((mobiltelefon.* | handy) & kein.netz)
	2.8 Education	((grundschul .* | gymnasium | .*schule.* | schulferien | schulkiosk | schulcampus | schulgebäude | schulmaterialien) & (abgeschalt.* | abgesackt.* | abgeschnitt.* | ausfäll.* | beeinträchtigt | berflutete | beschädig.* | betrifft.* | betroffen | demoliert | fall.* | entsorg.* | fall.* | fehlend.* | fielen | fortgespült | geflutet | gesperrt | gestört.* | geschlossen | leerzupumpen | problem.* | schäde.* | sperren | störung.* | überschwemmt | überspüllt.* | unterbroch.* | unter.wasser | übergelauf.* | überfüllt.* | vernichtet | verschoben | weggerissen | weggespült.* | zerbroch.* | zerlegt.* | zerstör.* | zusammengebrochen | zusammengeschlossen)) | (.*unterricht.* & (einschränkungen | containern))
	2.9 Child care	((kinderg.* | kindertagesstätte.* | kinderkripp.* | kita.* | pfarrkindergarten) & (abgeschalt.* | abgesackt.* | abgeschnitt.* | ausfäll.* | beeinträchtigt | berflutete | beschädig.* | betrifft.* | betroffen | demoliert | fall.* | entsorg.* | fall.* | fehlend.* | fielen | fortgespült | geflutet | gesperrt | gestört.* | geschlossen | problem.* | schäde.* | sperren | störung.* | überschwemmt | überspüllt.* | unterbroch.* | unter.wasser | übergelauf.* | überfüllt.* | vernichtet | verschoben | weggerissen | weggespült.* | zerbroch.* | zerlegt.* | zerstör.* | zusammengebrochen | zusammengeschlossen))

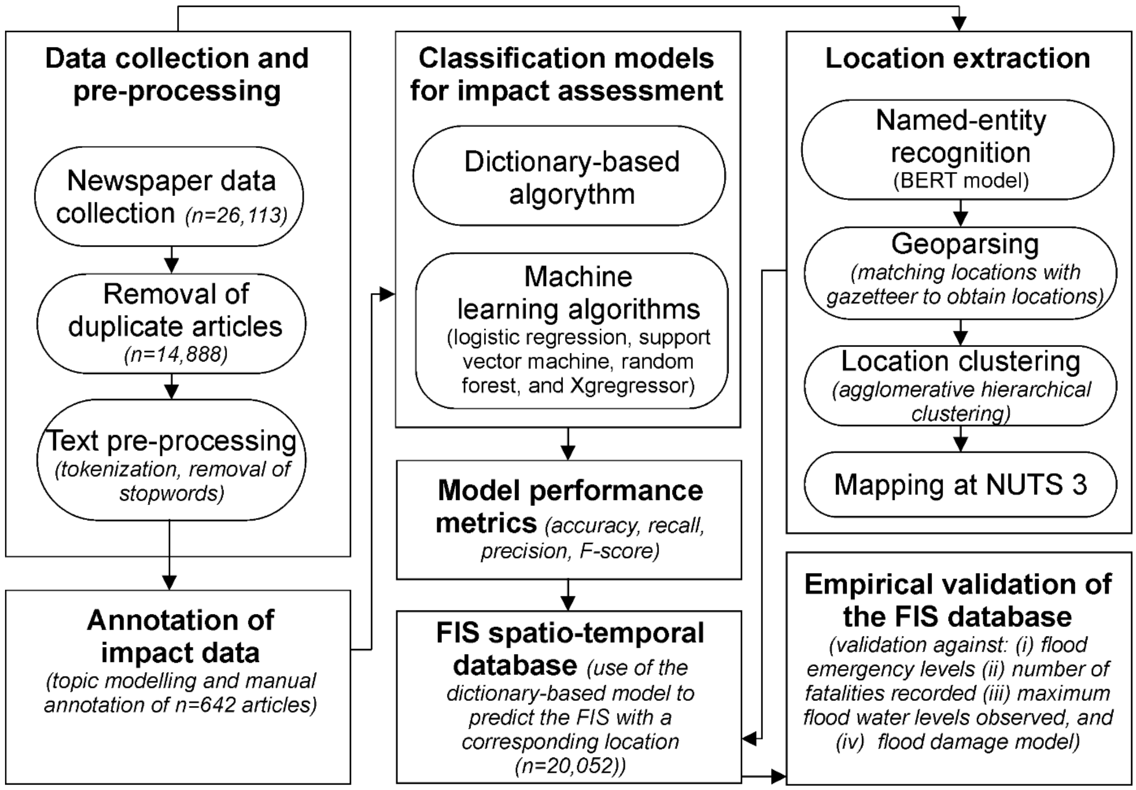
3. Economic	3.1 Need for economic support or monetary donations	bargeldspenden | geldspende.* | hilfspauschale| hilfsprogramm | finanzielle.hilfe.* | milliardenhilfe | staatliche.hilfen | hilfen.beantragen | hilfspauschale | spendensumme | spendenkonto | ((aufbauhilfen | gespendet.* | hilfe | hilfsbereitschaft | hochwasserhilfe | katastrophenhilfe | pauschal | sammelaktion | sammelt.* | soforthilf.* | spende.* | summe | summen | überwiesen | unterstütz.*) & (anträge | angewiesen | beantragen | euro.* | finanzielle | geld | gesammelt.* | hausratschäden))
4. Physical damage	4.1 Damage to buildings	(büro.* | einfamilienhäuser | eigentums | einrichtung | erdgeschoss | fachwerkhaus.* |.*gebäude.* | haus | hause | hauses | häus.* | häuser | heim | immobilien | keller | kellern | kirche | konferenzräume | kuche | materielle.* | mehrfamilienhäuser | privathaushalte | toilett.* | untergeschoss | wohngrundstücke | wohnhäuser | wohnung.* | zuhaus.*) & (abgerissen | abgesackt | betroffen.* | beschädigt.* | eingestürzt.* | einsturzgefährdet | flutet.* | getroffen | kaputt.* | geflutet.* | schaden | schäden.* | mitgerissen | nicht.mehr.stehen | schutt | überschwemmt | überflutet.* | unterspült | unter.wasser | verloren | vollgelaufenen | weggespült | weggeschwommen | zerstort.* | zerstört.* | zurückgebaut) | ((voll | volle) & keller)
	4.2 Damage to private assets	alles.verloren | sachbeschädigung | vieles.verloren | ((autos | autoanhänger.* | campingwag.* | fahrzeuge.* | grundstücke | hab.und.gut | kraftfahrzeug.* | kunstwerk | lkw | materiell | möbel | mobiliar | sachschäden | schränke | wohnmobile.*) & (abgesoffen | beschädigt | eingestürzt.* | enorm | gespült | getroffen | schaden | .*schäden.* | schrott | eingestürzt.* | geflutete | mitgerissen | überflutete | unbrauchbar | unter.wasser | weggespült.* | weggeschwemmt | verloren | zerstort.* | zerquetscht.*))
5. Economic sectors	5.1 Agriculture and livestock	ernteausfälle | ernteschäden | ernteeinbußen | ertragsausf.* | ((anbauflächen | bauern | ernte | ertrags | felder | feldern | getreid.* | heuernte | kartoffeln | landwirt.* | landwirtschaft | mais.* | raps | wein | weingütern | .*winzer.* | zuckerrüben) & (betroffene.* | euro | gefährdet | geschädigt | insolvenz | schaden | .*schäden.* | unter.wasser | vernichtet | verlust | überschwemmt | zerstört)) | ((heuernte | kühe | kuh | schafe | schafen | reh | rehe | rinde.* | schweine.* | tierbestände |vieh | ziege | ziegen) & (ertrunken | euro | futter.* | gefährdet | geschädigt | insolvenz | schaden | .*schäden.* | silageballen | unter.wasser | verlust | vorräten)) | futterspend.* | notfallwiese
	5.2 Industry and commerce	((bauunternehme.* | betriebe | geschaftsbetrieb | gewerbe.* |.*industrie.* | firmen.* | geschäft | geschäfte | geschäften | markt | handelskamm.* | sihk | versicherungswirtschaft | unternehmen | wirtschaft | wirtschaftsbetriebe) & (anträge | beschädigten | betroffen.* | euro | folgen | getroffen | geschlossen | insolvenz | schaden | *schäden.* | unterspülten | verlust | weggespült.* | zerstört.*))
	5.3 Tourism and gastronomy	((biergart.* | bäckerei | backstuben | café.* | camping.* | ferienhäuser |.*hotel.* | kneipe |.*museum.* | restaurant.* | gastronom.* | pizzeria | touri.*) & (folgen | insolvenz | shaden |.*schäden.* | unter.wasser | verlust | weggespült.* | zerstört.*))

633

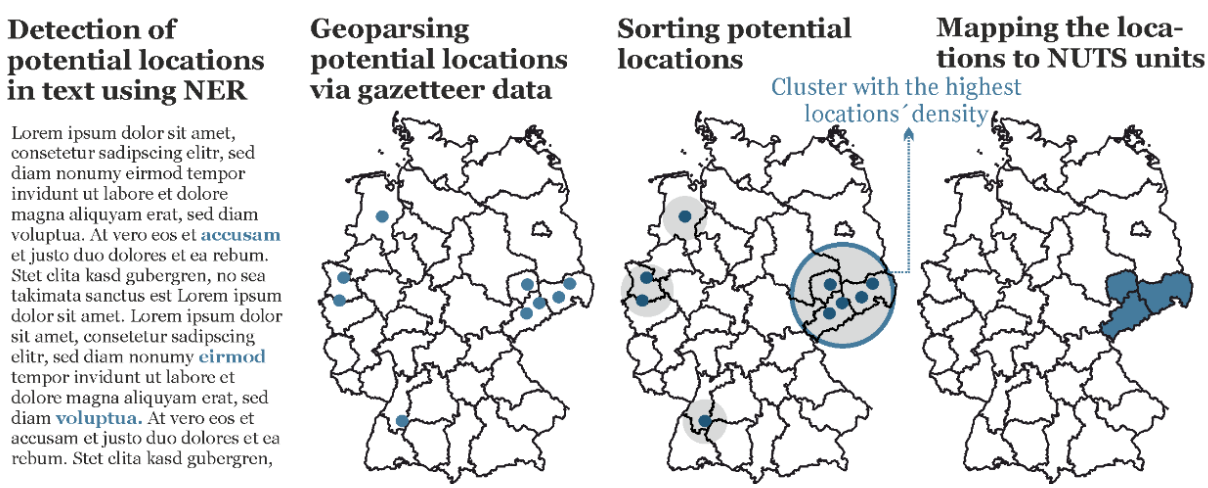
Supplementary Table S5. F-score for the flood impact statements (FIS) sub-classes according to different multilabel models.

FIS Sub-class	N of labelled data	XGboost	Random forest	Logistic regression	Support vector machine	Dictionary based
Agriculture & livestock	58	75.25	67.26	47.68	67.26	89.99
Buildings	207	84.87	56.87	40.37	64.80	93.83
Child care	55	83.55	47.25	47.25	52.14	95.50
Economic support	169	84.72	63.44	43.20	65.89	87.48
Education	55	81.77	51.44	46.96	51.44	91.83
Eletricity, gas & heating	95	86.21	48.86	45.61	51.93	93.48
Industry & commerce	61	75.76	51.78	47.11	51.78	92.17
IT & communication	50	68.35	54.35	47.97	54.35	88.99
Loss of income	53	88.44	48.11	48.11	48.11	93.43
Loss of lives	110	73.81	51.27	45.30	59.11	96.48
Mental health	56	80.70	51.78	47.11	59.35	96.48
Physical injuries	52	80.75	59.20	47.83	59.20	96.91
Private assets	67	79.19	55.52	48.25	61.84	88.27
Sewage system	53	73.68	52.93	47.54	52.93	82.81
Temporary displacement	55	48.11	48.25	48.25	48.25	80.56
Tourism	51	55.56	47.54	47.54	47.54	88.47
Transport & mobility	143	73.99	47.11	42.86	52.59	89.27
Waste & debris	73	66.66	47.40	47.40	47.40	92.50
Water contamination	50	75.25	47.83	47.83	53.84	97.96
Water supply	69	80.47	51.11	46.81	51.11	83.00

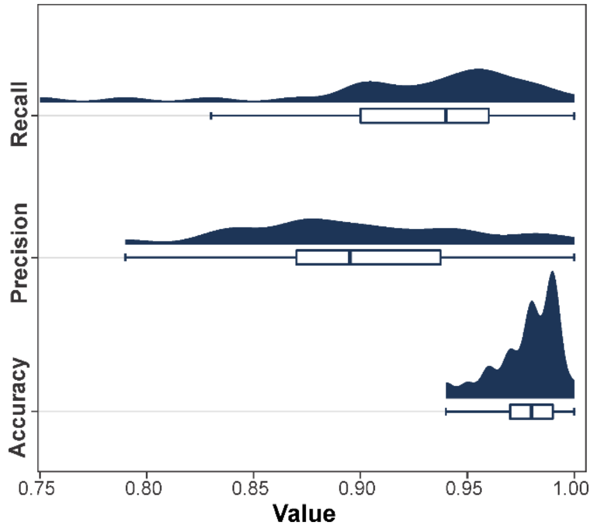
Supplementary Figure S1: Methodology overview, from data collection to empirical validation of the obtained FIS dataset



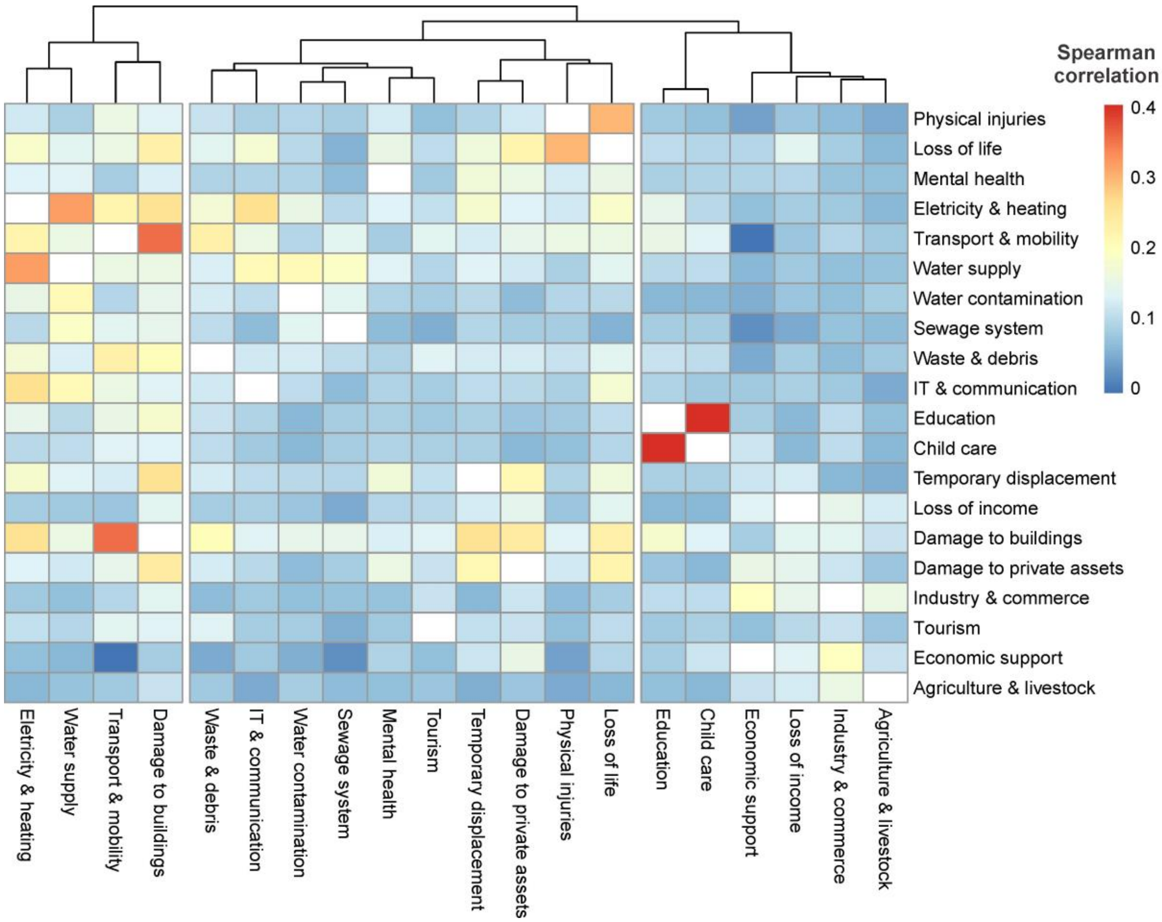
Supplementary Figure S2: Location extraction procedure. Adapted from Sodoge et al. (2022)



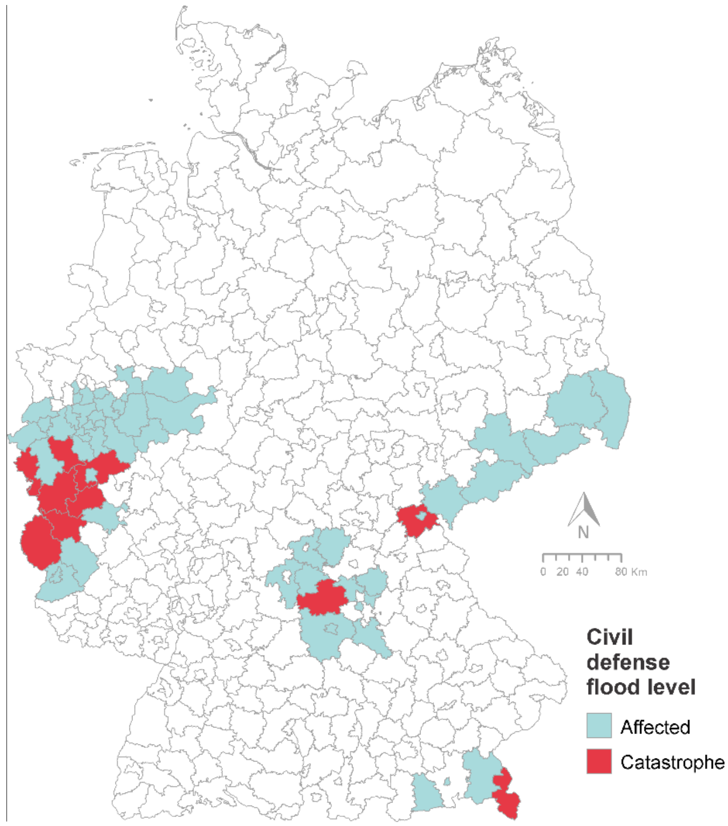
Supplementary Figure S3. Accuracy, precision and recall evaluation metrics for the socioeconomic impact classification model based on the dictionary-based approach. The model accuracy evaluates the overall proportion of correctly classified impacts, whereas the recall and precision assess the prevalence of false negatives and false positives.



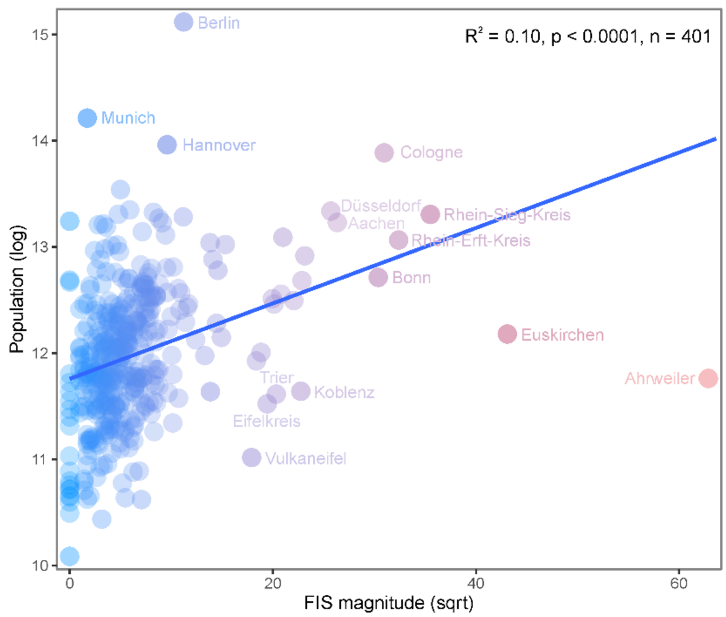
Supplementary Figure S4. Spearman correlation coefficients for the co-presence of FIS in the same newspaper article. The impact classes are grouped following a hierarchical clustering



Supplementary Figure S5. Civil Defense affected NUTS 3 regions between July and December 2021.



Supplementary Figure S6: Scatter plot depicting the FIS magnitude at the district level and the population size. Each circle represents a district. The coefficient of determination (R2) was determined to evaluate the fitness of the empirical model. Spearman correlation = 0.31, 0.23-0.41 CI, *p* = 0.0007.



Supplementary Figure S7: Number of FIS over time. The FIS were normalized by the number of articles in each month