

TOWARDS A SYSTEMATIC USE OF WEB-TEXT DATA TO SUPPORT GEOSPATIAL ANALYSIS OF MAJOR NATURAL DISASTER AND CRISIS EVENTS - EVIDENCE FROM THE AHR TAL 2021 FLOODING, GERMANY

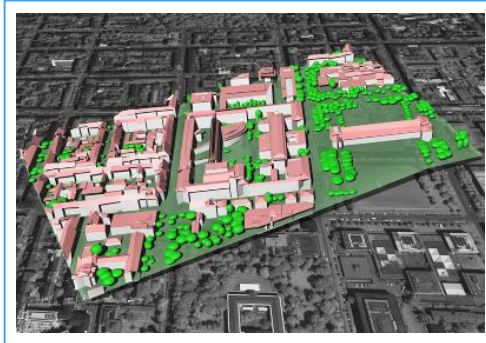
Vanessa Rittlinger, Sheikh Mastura Farzana, Xuke Hu, Hema Pandey, Stefan Voigt, Christian Geiß, Hannes Taubenböck

German Aerospace Center, Oberpfaffenhofen/Cologne/Jena, Germany

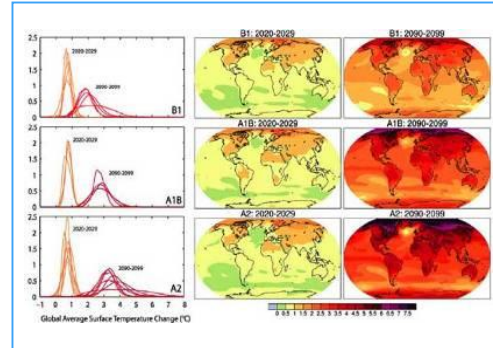
6th International Open Search Symposium – Munich/Garching October 2024



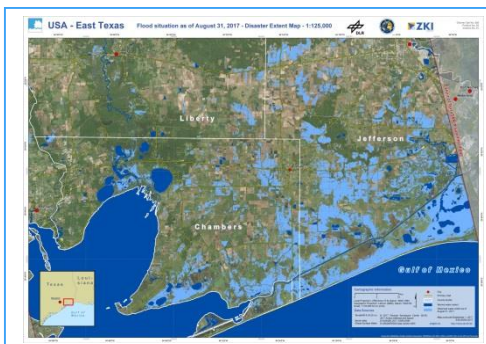
Topic examples



Civil engineering and urban planning



Global change research

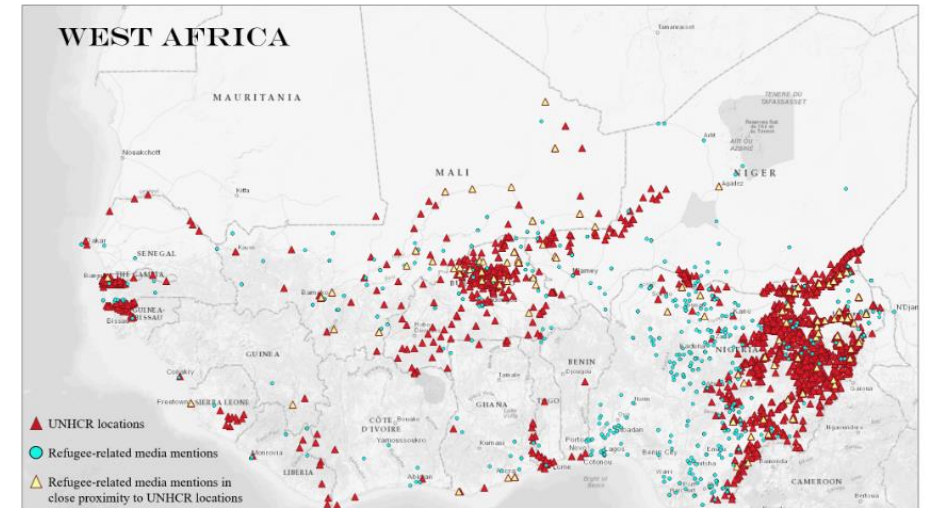


Safety and security

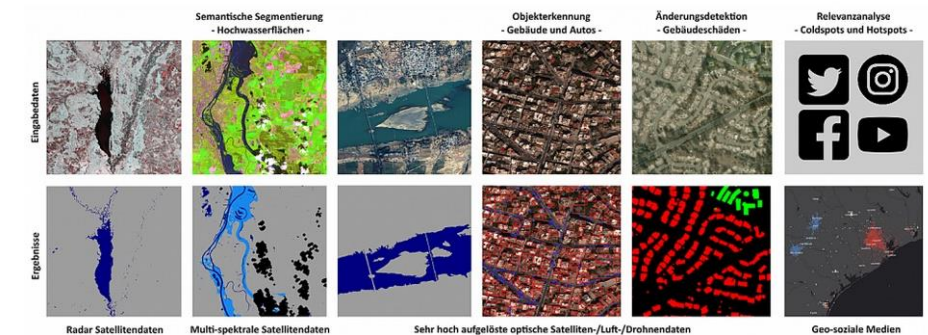


Sustainable development

Example use cases of NLP and EO data



Senaratne et al. 2023



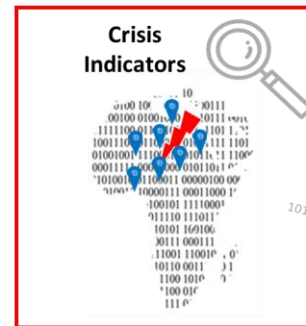
<https://crisis-prevention.de/kommunikation-it/entscheidungsunterstuetzung-im-katastrophenschutz.html>, last accessed 08.10.2024

OpenSearch@DLR & use of OpenWebSearch.eu Index data at DLR Earth Observation Center



Topic relevant project focus:

- Scientific search and „geoparsing“ of web-text data at scale
- Statistical analysis of georeferenced web-text data for combined analysis with satellite based Earth Observation data
- Particular use case: crisis- and disaster situational awareness (ex post as well as in near real-time)



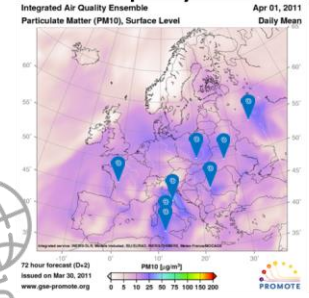
Data and Citizen Science:



Analyzing Global Urbanization:



Airquality



...and many more!



Motivation

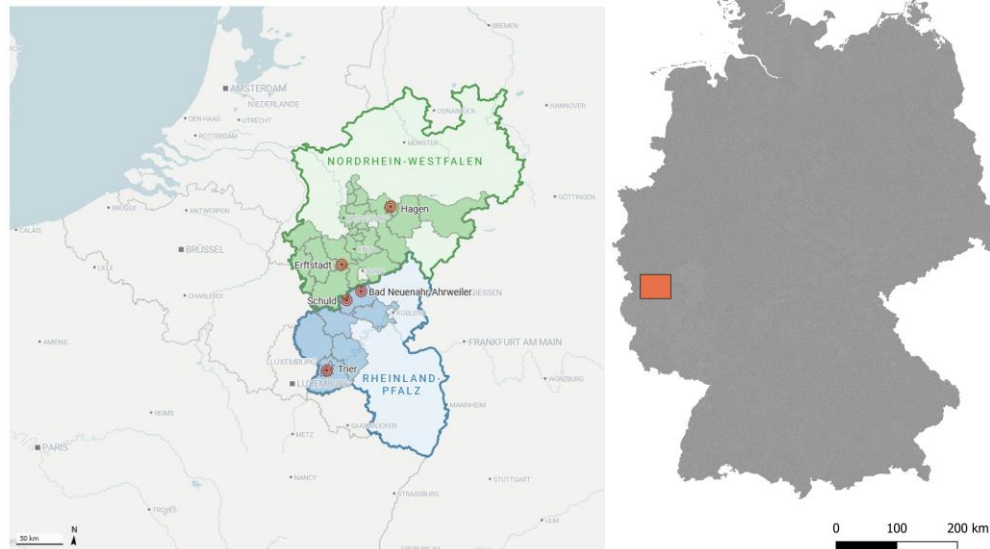


- Extraction of relevant **geo-risk related data** from web-text data
- **Derive detailed geolocated information** on different georisk parameters
- User crawled and indexed web-text data to reconstruct **temporal** and **spatial** aspects of disaster events
- **Synergistically analyze web-text derived geo-information with satellite** based data and other sources of geospatial data for better understanding of natural **disasters** and **crisis** situations.

Ahrtal 2021 flooding event

Event details:

- Flood event in 2021 in western Germany (riverine flood)
- Caused through heavy rain events from low pressure area „Bernd“
- Saturated soil from precipitation on the days before



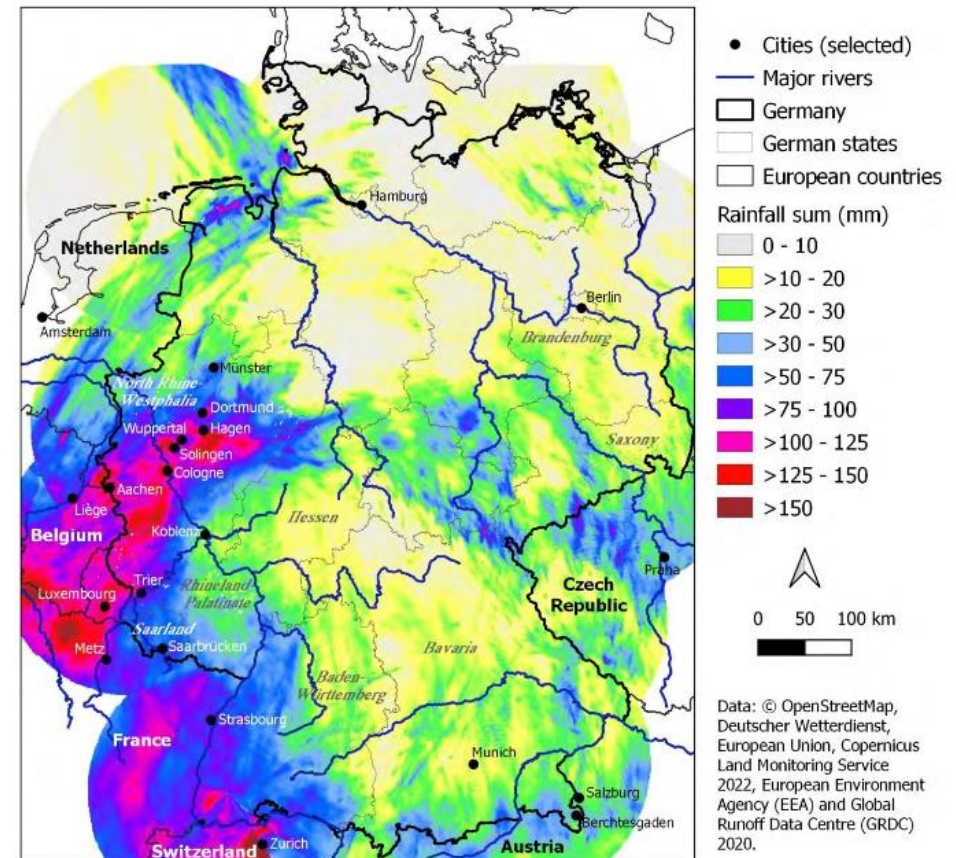
Grafik: bpb - Quelle: Bundesamt für Bevölkerungsschutz und Katastrophenhilfe (BBK)

<https://www.bpb.de/kurz-knapp/hintergrund-aktuell/337277/jahrhunderthochwasser-2021-in-deutschland/> (accessed 24.05.2024)

Rittlinger et al., #osym2024, October 2024, Munich/Garching

Consequences:

- 33 billion Euro of economic loss
- Human casualties

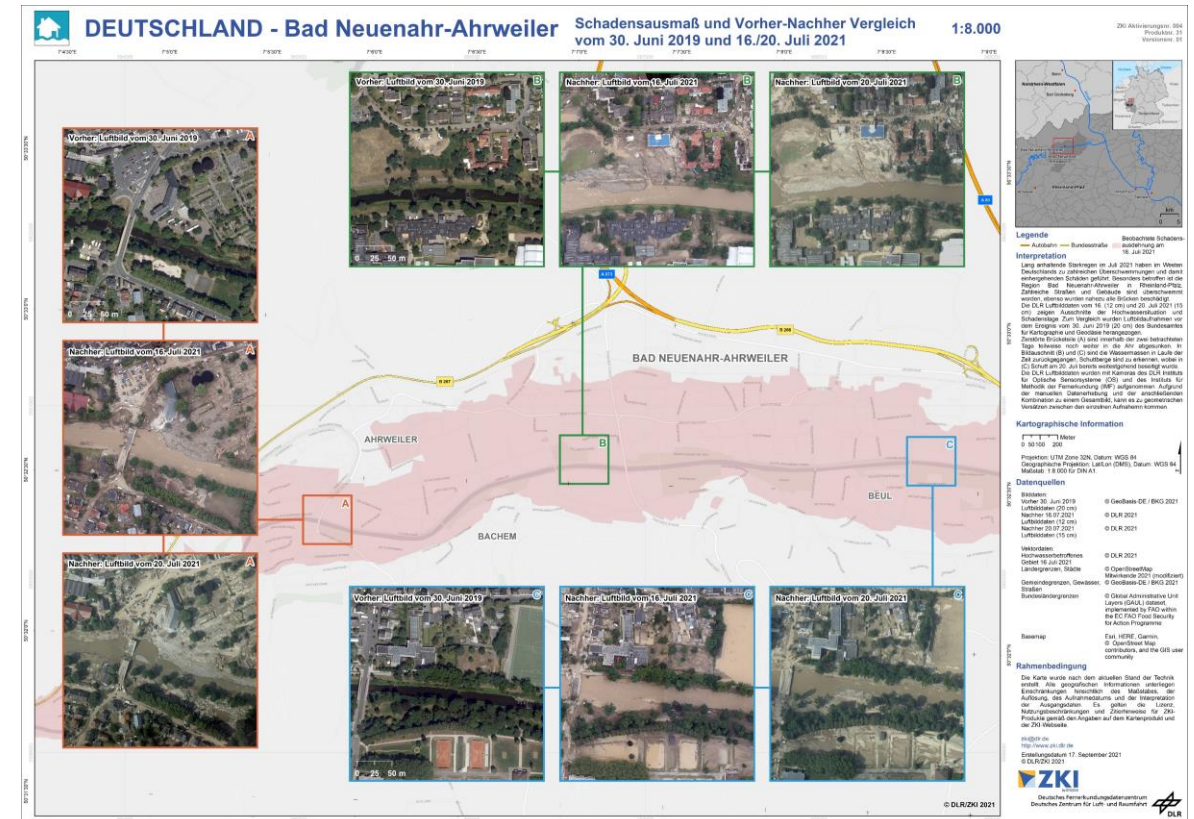


Kron et al. 2022; Rackelmann et al. 2023

Ahrtal 2021 flooding event – Remote sensing based emergency mapping and analysis



<https://activations.zki.dlr.de/de/activations/items/ACT152/>, last accessed 08.10.2024



<https://activations.zki.dlr.de/de/activations/items/ACT152/>, last accessed 08.10.2024

Ahrtal 2021 flooding event – Information on the web...



Scientific projects

The screenshot shows the website 'BER KAHR' with a navigation menu including 'BER KAHR', 'PROJEKTPARTNER', 'AKTUELLES', 'MEDIATHEK', 'INFORMATIONEN', 'PUBLIKATIONEN | PRESSE', 'IMPRESSUM', and 'DATENSCHUTZ'. The main content area features a header image of a flooded river and a sub-header 'ber: Startseite / Über KAHR'. Below this, there is a logo for 'KLIMA ANPASSUNG HOCHWASSER RESILIENZ' and a text block describing the project's goals: 'Das Hochwasser im Juli 2021 hat zu einer extremen Betroffenheit an Erft, Inde, Vichtbach und Wupper in Nordrhein-Westfalen (NRW) sowie an der Ahr in Rheinland-Pfalz (RLP) geführt. KAHR begleitet den (Wieder-)aufbauprozess und leistet einen wissenschaftlichen Beitrag zum Hochwasserrisikomanagement nach der Flutkatastrophe, um betroffene Regionen resilienter zu gestalten.' A list of planned activities follows, including 'Überarbeitung der Hochwassergefahrenkarten', 'Weiterentwicklung von Hochwasser-/Starkregenrisikokonzepten', and 'Darstellen von Möglichkeiten und Hemmnissen für resilienten Wiederaufbau und Innovation'.

<https://hochwasser-kahr.de/index.php/de/>, last accessed 04.10.2024

Event facts and reanalysis

The screenshot shows the website 'kurz&knapp' with a navigation menu including 'inemen', 'Mediathek', 'shop', 'Lernen', 'veranstaltungen', 'kurz&knapp', 'die opp', 'menü', and 'EN'. The main content area features a header image of a flooded river and a sub-header 'kurz&knapp Hintergrund aktuell Zahlen & Fakten Lexika Dosis Politik'. Below this, there is a title 'Nach der Flut an der Ahr 2021 Wiederaufbau und Aufarbeitung' and a text block: 'Über 180 Menschen starben bei der Flut in Rheinland-Pfalz und Nordrhein-Westfalen im Juli 2021. Ein Überblick über den Stand des Wiederaufbaus und die politische Aufarbeitung zwei Jahre danach.' The page also includes a 'Redaktion' section with the date '10.07.2023 / 6 Minuten zu lesen' and social media sharing icons.

<https://www.bpb.de/kurz-knapp/hintergrund-aktuell/522893/nach-der-flut-an-der-ahr-2021/>, accessed 04.10.2024

News and media

The screenshot shows a WDR reportage titled 'Ahrtal unter Wasser' with a subtitle 'Chronik einer Katastrophe'. The main image is an aerial view of a flooded town. The navigation menu includes 'Chronik Hochwasser im Ahrtal', 'Vor der Flut', 'Hochwasser-Karte Ahrtal', 'Frühe Warnungen', 'Ahrweiler Krisenstab tagt', and 'Ahr-Pegel'. The text 'Ahrtal unter Wasser' is prominently displayed in white over the image, with 'Chronik einer Katastrophe' below it.

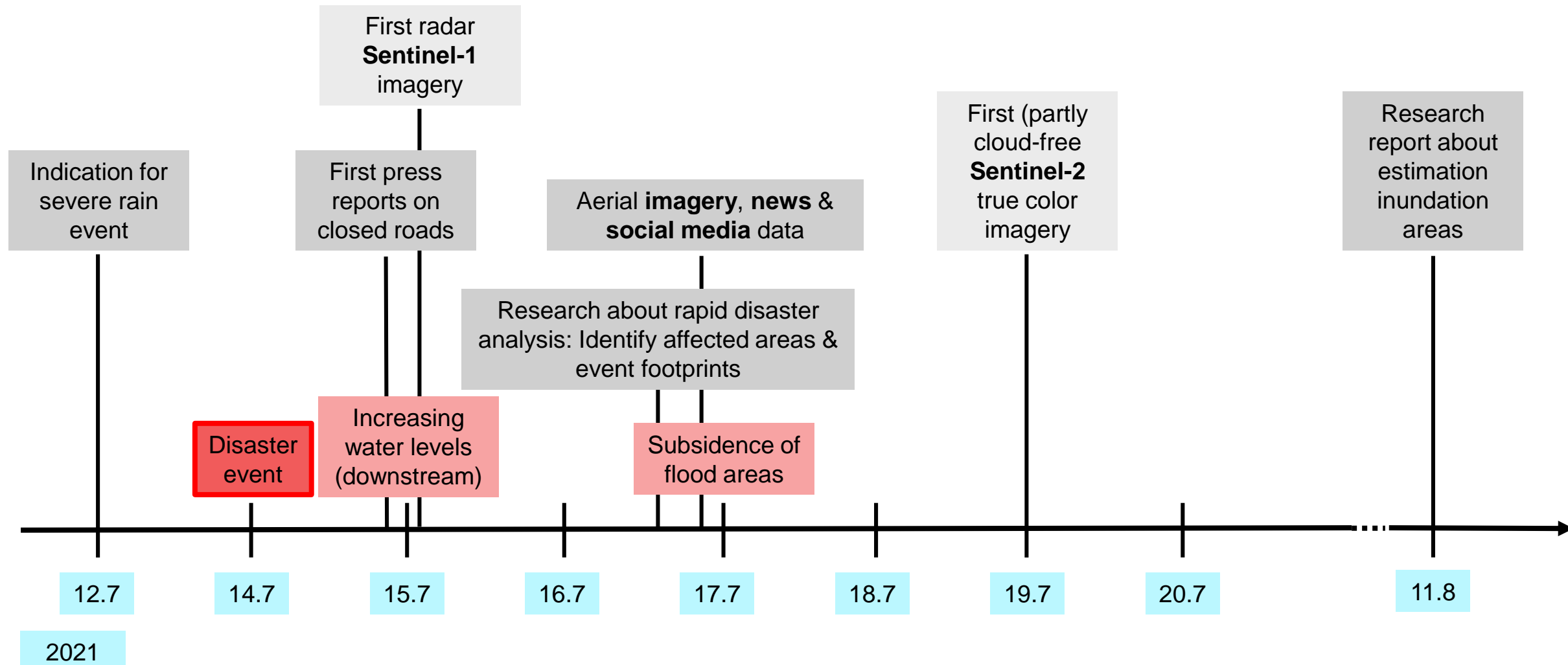
<https://reportage.wdr.de/chronik-ahrtal-hochwasser-katastrophe#vor-der-flut>, accessed 04.10.2024

Political statements

The screenshot shows the website 'Rheinland-Pfalz' with a navigation menu including 'Unser Land', 'Themen', 'Ministerpräsidentin', 'Regierung', and 'Service'. The main content area features a header image of a flooded river and a sub-header 'Rheinland-Pfalz DIE LANDERREGIERUNG'. Below this, there is a photograph of three people standing on a bridge over a flooded river. The text 'Rheinland-Pfalz' is prominently displayed in white over the image, with 'DIE LANDERREGIERUNG' below it.

<https://www.rlp.de/ministerium/unser-land/wiederaufbau-nach-der-flut>, accessed 04.10.2024

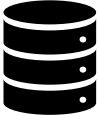
Event timeline and data availability



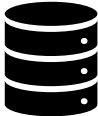
From web-text data to geo-risk points



Datasets



Focused crawl

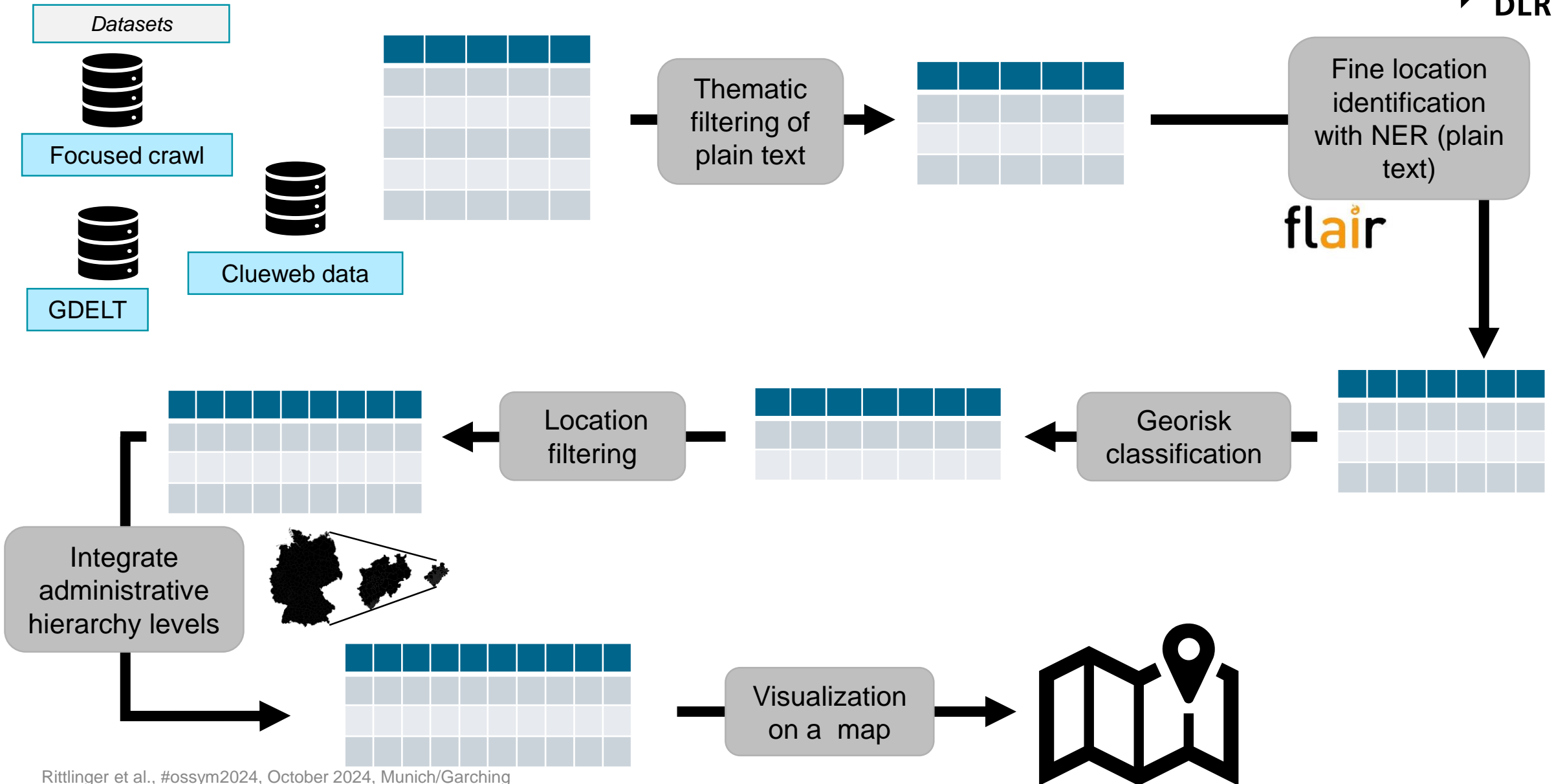


GDELT



Clueweb data

From web-text data to geo-risk points



Different web-text data sources (overview)



Data derived by OpenSearch@DLR project

Complimentary data

FOCUSED CRAWL

- Crawled websites with focus on flood and Ahrtal flooding event related weblinks
- Coarse geocoding applied
- Information and parameters set within project

GDELT DATA

- Knowledge graph data from 01.07.2021 – 31.12.2022
- Data from news websites
- Topics including flood
- Study area bounding box
- Other examples of disaster analysis (Owuor et al. 2020)

The GDELT Project. <https://www.gdeltproject.org/about.html>, last accessed 04.10.2024

CLUEWEB DATA

- Clueweb 2022 crawled web
- Language: German
- Filtered for 235 municipality names in Ahr valley
- For each location max. 200 websites returned

Clueweb dataset <https://lemurproject.org/about.php>, accessed 04.10.2024

Different web-text data sources (overview)



Data derived by OpenSearch@DLR project

Complimentary data

FOCUSED CRAWL

- Crawled websites with focus on flood and Ahrtal flooding event related weblinks
- Coarse geocoding applied
- Information and parameters set within project

GDELT DATA

- Knowledge graph data from 01.07.2021 – 31.12.2022
- Data from news websites
- Topics including flood
- Study area bounding box
- Other examples of disaster analysis (Owuor et al. 2020)

The GDELT Project. <https://www.gdeltproject.org/about.html>, last accessed 04.10.2024

CLUEWEB DATA

- Clueweb 2022 crawled web
- Language: German
- Filtered for 235 municipality names in Ahr valley
- For each location max. 200 websites returned

Clueweb dataset <https://lemurproject.org/about.php>, accessed 04.10.2024

Approach:

- Plain text of websites in German
- Websites data after July 2021
- Location filter of study area
- Different processing steps due to different parameter information

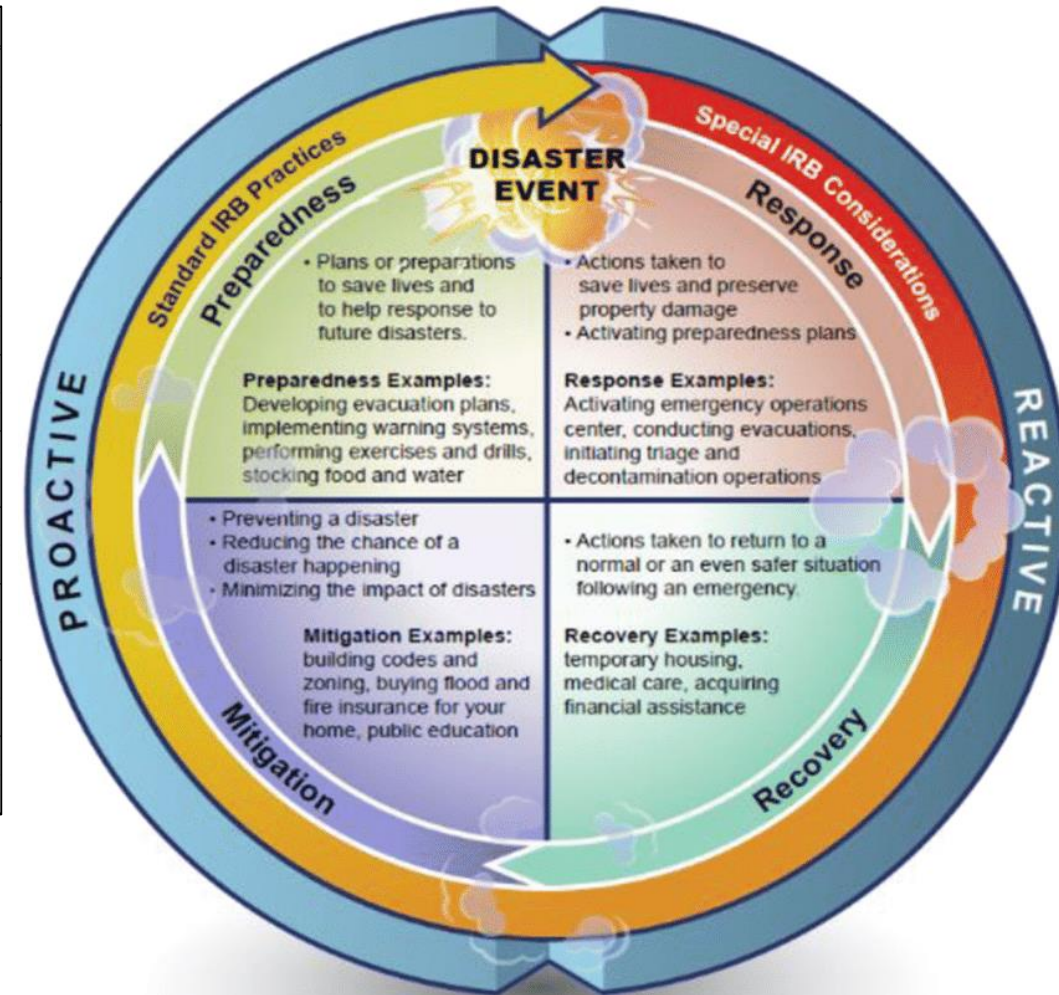
Method: Thematic filtering

Main topic	Class	Subclass	Words
Flut [Fluten]	Synonyms	Synonyms	Hochwasser
			Überschwemmung
	Preparedness	Risk assessment ->prevention	Hochwassergefahr
			Naturgefahr
			Prediction / Early Warning
	Disaster/ Impact	Disaster	Starkes Gewitter
			Extremwetterereigniss
			Flutkatastrophe
	Response	Impact / Damage infrastructure	Starkregen
			[zerstörtes] Haus
		Impact / Human loss	[überflutete] Keller
			Tod
		Emergency Management (Personen/ Organisationen)	Vermisst
			Rettungskräfte
Emergency Management (Maßnahmen)		Einsatzkräfte / Fahrzeuge	
		Hilfsangebote	
Recovery	Reconstruction	Einsatzschwerpunkt	
		Wiederaufbau	
Prevention	Mitigation	Übergangsunterkunft	
		Hochwasserschutz	
			Katastrophenschutz



2 loop filtering process:

- Corase filtering: 84 words (Synonyms & Preparedness categories)
- Detailed filtering: 138 words (other categories)

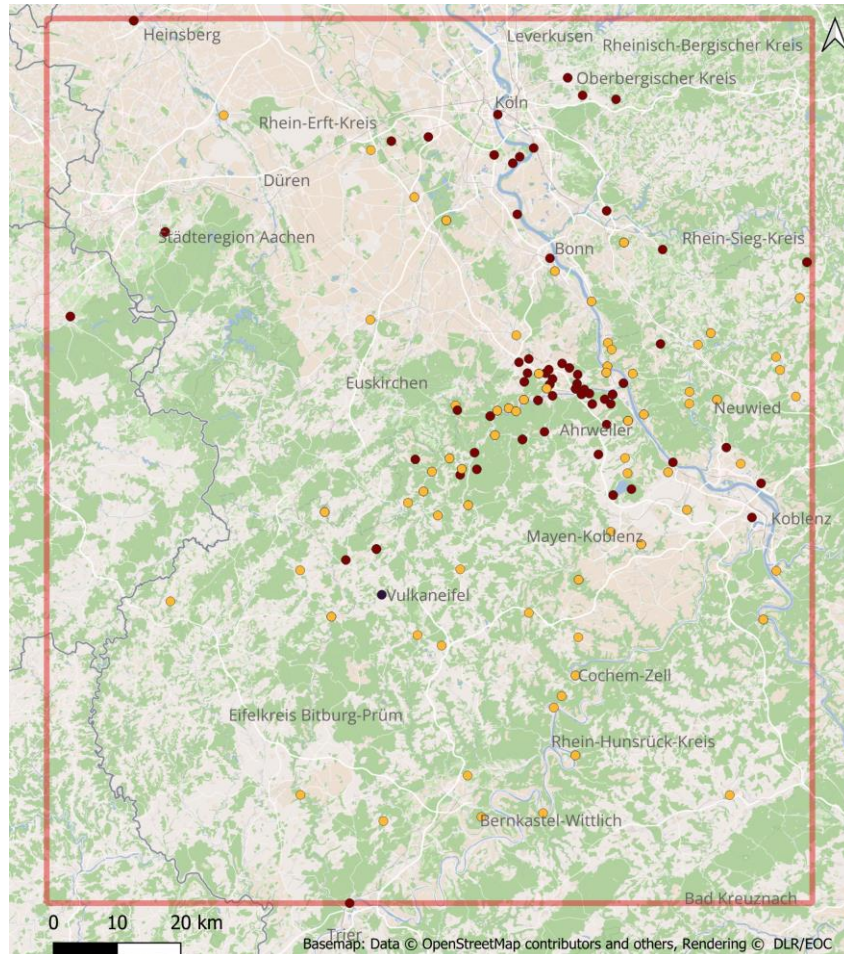


First results: First approach of plotting data from web-text



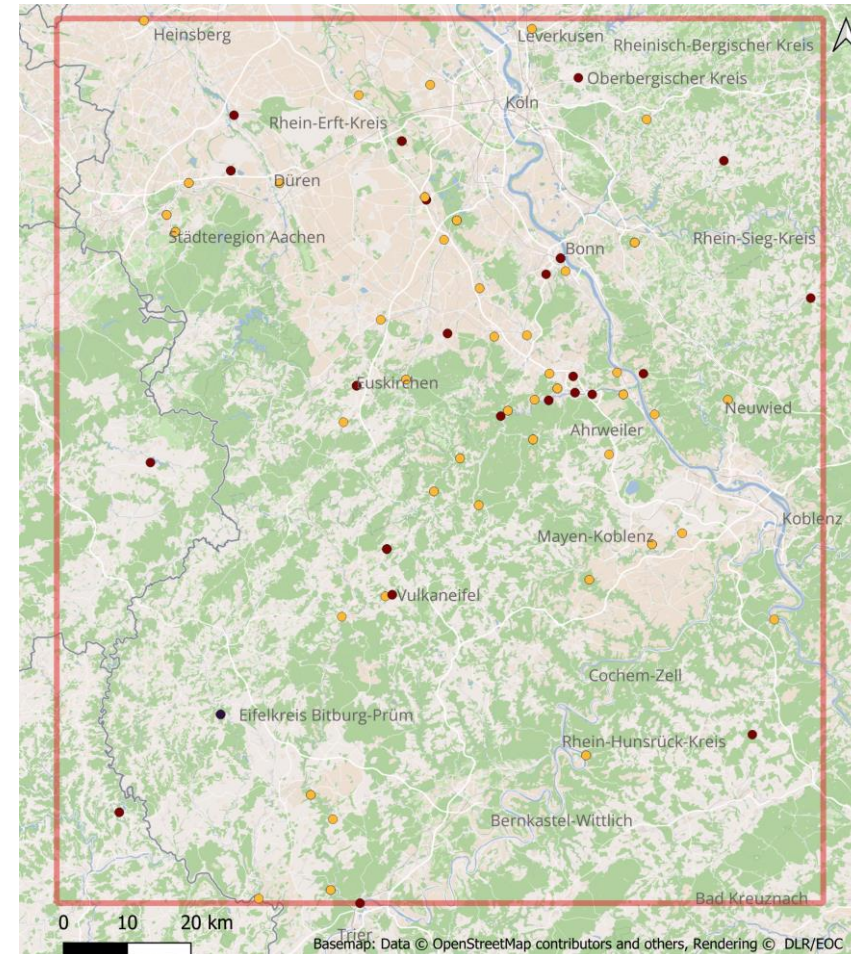
Focused crawl

2720 websites - 377 locations



Clueweb

3145 websites - 830 locations



- Administrative level 2
- Administrative level 3
- Administrative level 4
- No administrative level added
- Study area extent

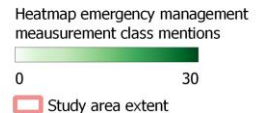
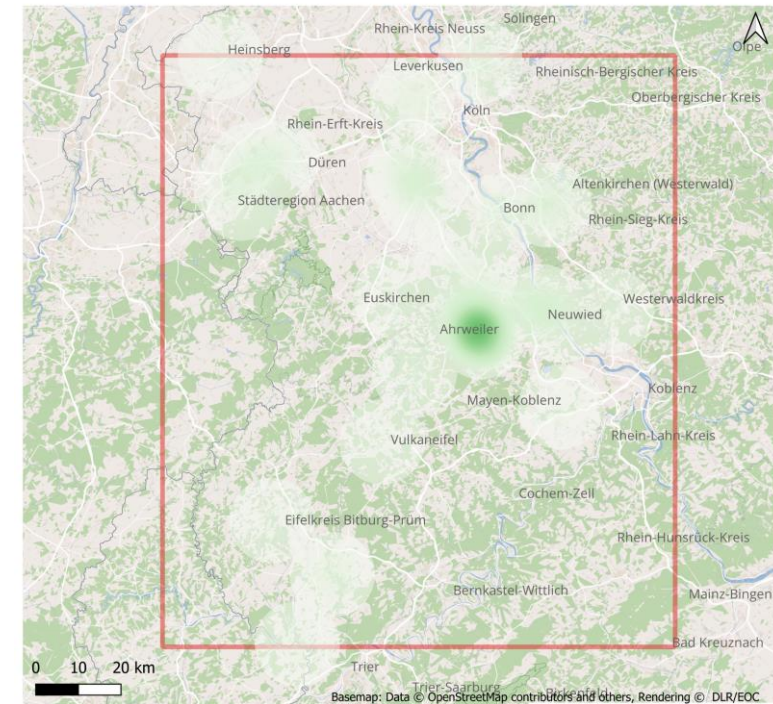
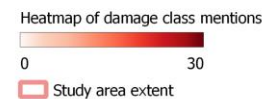
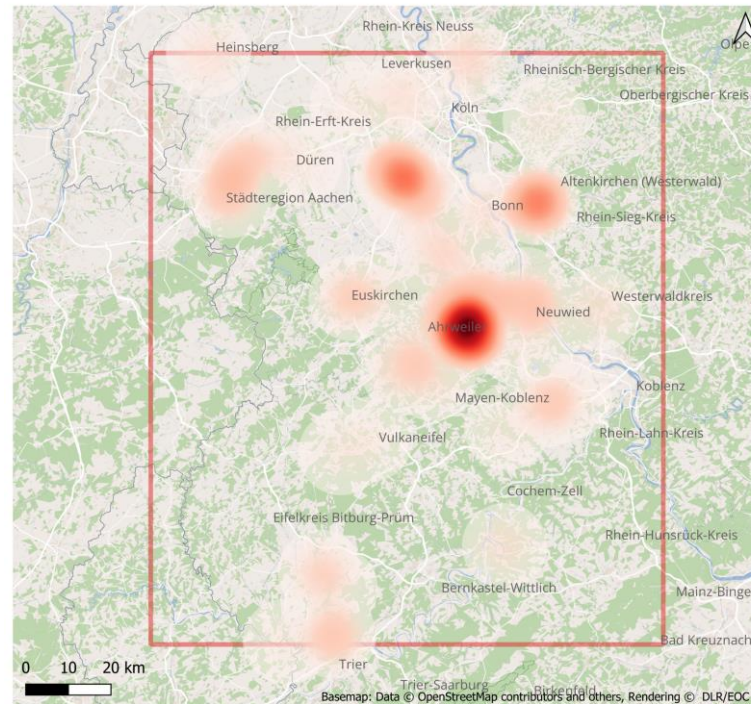
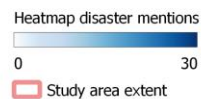
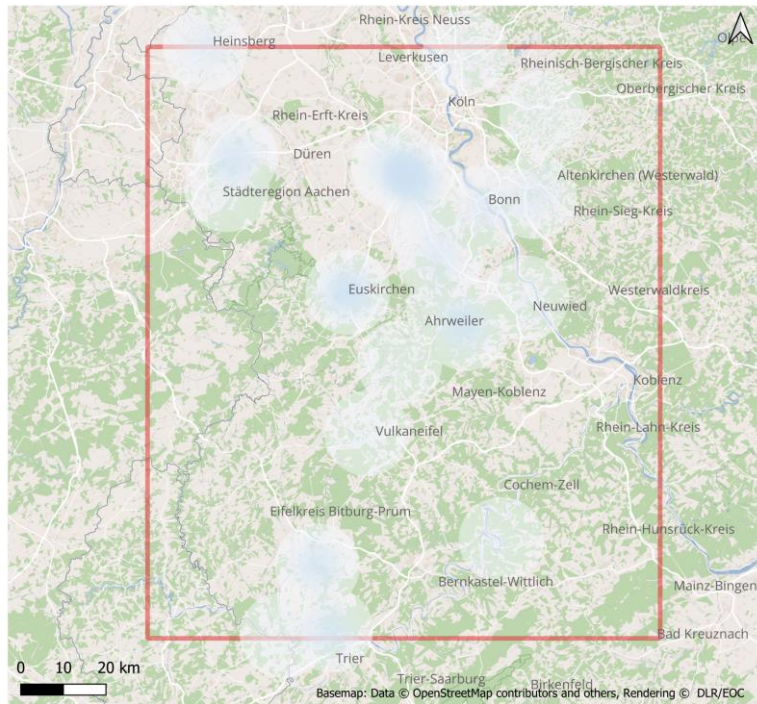
First results: Heatmaps of spatial and thematic information from focused crawl web-data for 2021 Ahrtal flood



Class „Disaster“

Class „Damage“

Class „Emergency management“



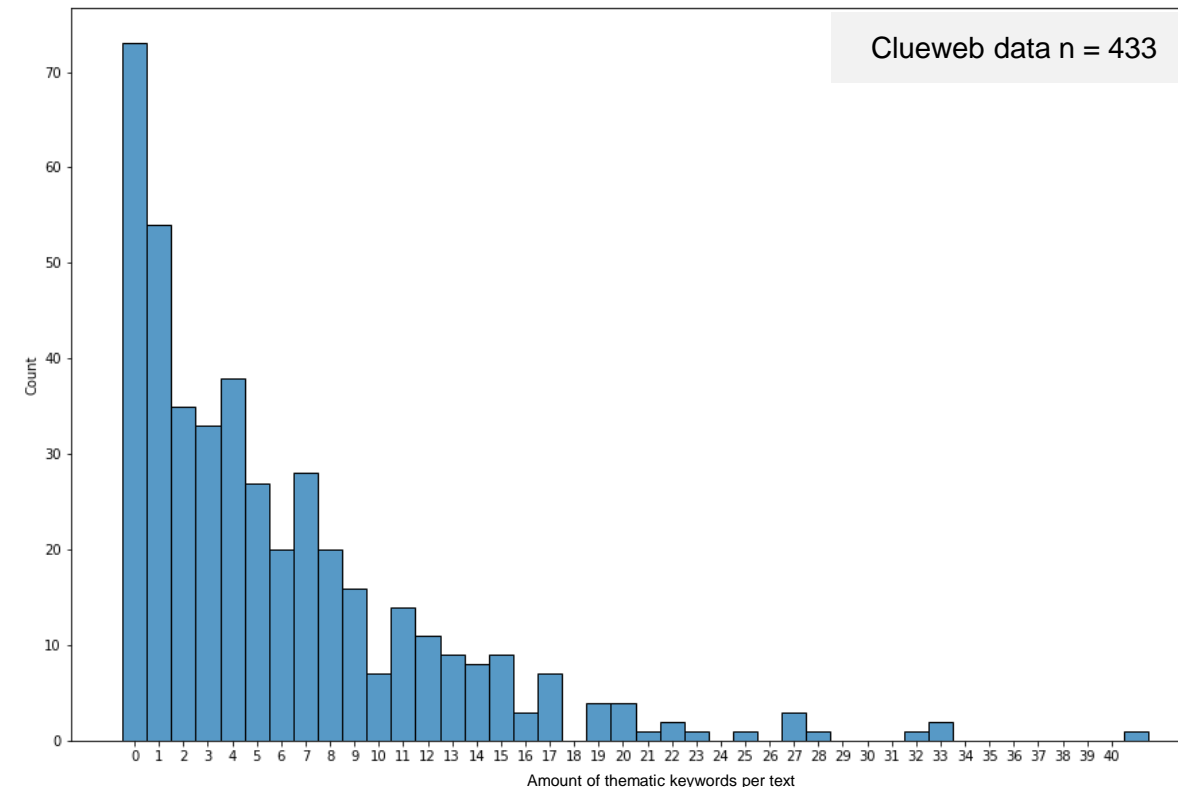
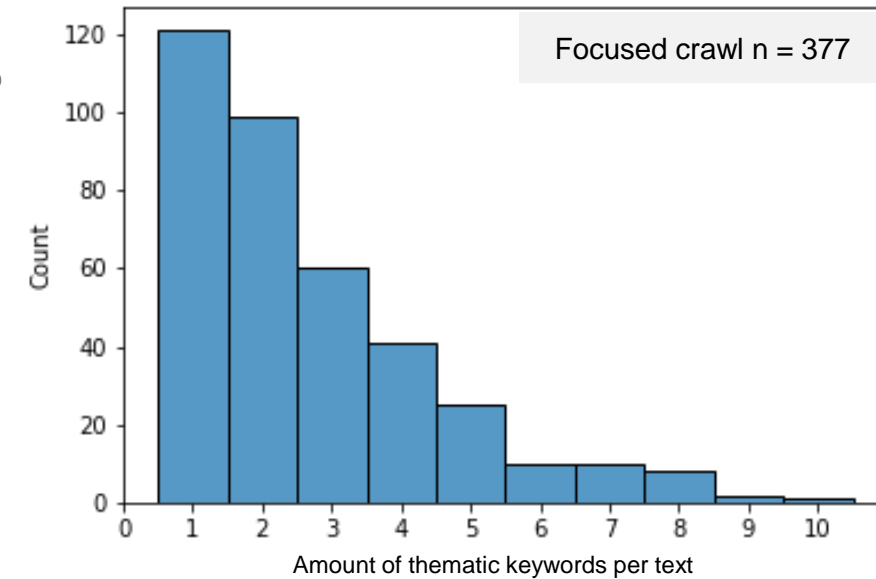
First results: Conclusions and Challenges



- Text and information contained in data sets differs significantly
- Temporal information is difficult to assess in web-text data
- Web-text data and text varies due to largely varying website types
- Better thematic filtering approach needed: e.g. unsupervised topic modeling
- Linking location and georisk class correctly

First results: Conclusions and Challenges

- Text and information contained in data sets differs significantly
- Temporal information is difficult to assess in web-text data
- Web-text data and text varies due to largely varying website types
- Better thematic filtering approach needed: e.g. unsupervised topic modeling
- Linking location and georisk class correctly



Upcoming work



- Application of advanced ML/ DL methods in NLP for the thematic disaster information extraction, including topic modeling approaches in the coarse filtering steps
- Improvement of the fine grained location information in combination of the retrieved geo-risk related information
- Assessment of spatio-temporal dynamics of the disaster
- Combination of web-text derived disaster information with other geospatial data sources such as satellite imagery to improve situational understanding

Any questions and discussion?



Literature references



- Kahle, M., Kempf, M., Martin, B., & Glaser, R. (2022). Classifying the 2021 'Ahrtal' flood event using hermeneutic interpretation, natural language processing, and instrumental data analyses. *Environmental Research Communications*, 4(5), 051002. <https://doi.org/10.1088/2515-7620/ac6657>.
- Kron, W., Bell, R., Thiebes, B., & Thieken, A. (2022). *The July 2021 flood disaster in Germany* (S. 12–44).
- Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J. E., Ehmele, F., Feldmann, H., Franca, M. J., Gattke, C., Hundhausen, M., Knippertz, P., Küpfer, K., Mühr, B., Pinto, J. G., Quinting, J., Schäfer, A. M., Scheibel, M., Seidel, F., & Wisotzky, C. (2022). *A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe. Part 1: Event description and analysis*. <https://doi.org/10.5194/nhess-2022-137>.
- Owuor, I., Hochmair, H. H., & Cvetojevic, S. (2020). Tracking Hurricane Dorian in GDELT and Twitter. *AGILE: GIScience Series*, 1, 1–18. <https://doi.org/10.5194/agile-giss-1-19-2020>.
- Pakenham, Joan & Rosselli, Richard & Fothergill, Alice & Slutsman, Julia & Ramsey, Steve & Hall, Janet & Miller, Aubrey. (2021). Institutional Review Board Preparedness for Disaster Research: a Practical Approach. *Current Environmental Health Reports*. 8. 10.1007/s40572-021-00311-x.
- Rackelmann, F., Sebesvari, Z., & Bell, R. (2023). Synergies and trade-offs in the management objectives forest health and flood risk reduction. *Frontiers in Forests and Global Change*, 6. <https://doi.org/10.3389/ffgc.2023.1208032>.
- Senaratne, H., Mühlbauer, M., Kiefl, R., Cárdenas, A., Prathapan, L., Riedlinger, T., Biewer, C., & Taubenböck, H. (2023). The Unseen—An Investigative Analysis of Thematic and Spatial Coverage of News on the Ongoing Refugee Crisis in West Africa. *ISPRS International Journal of Geo-Information*, 12(4), 175. <https://doi.org/10.3390/ijgi12040175>.