

RESILOC privacy notice regarding the research activity 'Social media analysis to be used in Gorizia Trial'

The Resilient Europe & Societies by Innovating Local Communities (RESILOC) project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 833671. The project's objective is to increase Europe's resilience to crisis and disasters. Resilience is defined by the United Nations as "the ability to resist, absorb and accommodate to the effects of a hazard, in a timely and efficient manner". Thus, resilient communities are those in which their citizens, environment, businesses, and infrastructures have the capacity to withstand, adapt, and recover in a timely manner from any kind of hazards they face, either planned or unplanned. In recent years, efforts have been spent to tackle resilience and there is, still, a long path forward in defining an EU valid and sound approach to the problem. RESILOC aims at studying and implementing a holistic framework of studies, methods and software instruments that combines the physical with the less tangible aspects associated with human behaviour.

You can find more about us here: <https://www.resilocproject.eu/>

The social media analysis, for which we issue this privacy notice, is an activity based on aggregated information from tweets. It is an experimental activity with Twitter data that aims to capture specific information, such as the volume of tweets and their sentiment (positive, negative or neutral) on topics that are predefined by the Community of Gorizia in its role in the RESILOC community.

Results from this analysis are going to be used by the Local Resilience Team of Gorizia as the main users of the RESILOC tools. The results will support their decision on how to set up different proxies for different indicators along several dimensions of resilience and to estimate whether the aggregated data from tweets is potentially useful as a proxy in the resilience assessment of their community.

Tweets that are used are public on the Twitter platform and access to them is granted to the RESILOC consortium (particularly partner JSI) for the purposes of research use in the project based on an agreement between Twitter and the consortium.

For the Gorizia Trials, the query used is "Gorizia"; during the activity, the query can be updated at a later stage to focus on "Nova Gorica". For the query, textual contents of tweets are extracted to be used in the project. Then, tweets are filtered according to their language and a sentiment label (positive, negative, neutral) is prescribed by sentiment analysis algorithms (for additional information on the algorithms, please consult the end of this section). Text of the tweets is used also for automated detection of topics in collections of tweet texts. Yearly and monthly aggregates of tweet sentiment, volume and frequent tokens (words) are calculated which represent the results of the analysis. Only these aggregated results are presented to the participants of the pilots via password protected website.

In the trial, we use only the aggregated data that is to be provided to the Gorizia trial participants. The analysis of individual tweets is done beforehand on separate offline computers at JSI and individual tweets are discarded once their contents is no longer needed for the analyses.

All analyses will be ended by the end of the project. However, the aggregated data will remain available to project partners for the purposes of the project and its research aims.

Additional information on the sentiment analysis algorithms:

Sentiment analysis employs a sentiment classifier - an algorithm that takes text as input and outputs its classification into one of the sentiment classes (positive, negative, neutral). These algorithms are trained on large amounts of texts with (manually) prescribed sentiment classes. An example of such a classifier would represent a text item as a collection of words. Based on the training data, each word found in the training data would, for example, be assigned a weight or vote for a given class (such as positive or negative). When a new item needs to be classified, votes of consisting words would be used to assign a class for the entire new text item.

Modern classifiers (such as the ones used in RESILOC) work mostly in the same way but use somewhat more elaborate representations of text items. The first of two classifiers that we used was a logistic regression classifier trained on high-dimensional vector representations that included weighted words, couples of consecutive words, 4-character sequences and emoji characteristics as representation elements (such as the approach presented in this paper: http://ceur-ws.org/Vol-1866/paper_78.pdf). The second classifier is the FEEL-IT sentiment classifier for Italian that represents words or subword series of characters in a high dimensional vector space, where they are placed according to training data in a way that proximity in the vector space corresponds with the usual neighbourhood (context) of the word or subword (details on this approach are available in this paper: <https://aclanthology.org/2021.wassa-1.8.pdf>).

There are no foreseeable risks associated with the personal data of users whose tweets are used in the analysis. The data used in the trial represent aggregates of tweets, therefore, it does not include individual tweets. In this way, the data will be anonymised, or at least pseudonymised, and care will be taken that any

personal information that could be used to identify users who tweeted will be removed if detected in the aggregates.

However, there are several analysis-related risks that need to be considered. The tool used to assess the sentiments is still in the research phase. This research exercise will support its improvement. Moreover, the data may be insufficient, so the sentiment assessments may not be entirely accurate. On the other hand, tweets reflect personal opinions (opinion segments) on current issues of global or local nature.

The above may reflect on the interpretation of the results by the Local Resilience Team. However, RESILOC researchers are aware of this and the research team will instruct the Local Resilience Team to be mindful of it as well.

Privacy and confidentiality have been carefully considered in the RESILOC project to meet legal requirements, as well as ethical considerations. Data collected throughout the project activities will be used to develop the RESILOC project objective, which is to increase Europe's resilience to crises and disasters. Only data that is necessary for the development of the project will be collected.

As data subjects, the tweets' owners are entitled to the data subject rights under GDPR. These include:

- obtaining access to their personal data processed by us, if any;
- obtaining the rectification of inaccurate data or the completion of their personal data, if any has been collected by RESILOC during this exercise;
- obtaining the erasure of personal data saved by us (right to be forgotten), if any – this right cannot be exercised for analysis that is already complete;
- requesting restriction of processing of their personal data, if any personal data has been collected;
- objecting to the personal data processing, if any personal data has been collected;
- requesting not to be subject to a decision based solely on automated processing, including profiling;
- receiving their personal data that we have collected (if any), in a structured, commonly used and machine-readable format;
- transmitting those data to another entity;
- lodging a complaint with a supervisory authority or the competent courts of the Member States.

Only authorised persons assigned through internal project processes will have access to personal data. RESILOC follows the data minimization principle (Article 5 GDPR). This principle consists of collecting only the necessary amount of data and using it to the minimum extent. Therefore, data that is no longer required to develop project activities will be deleted.

For this particular research activity, we access and process the following potentially personal data from tweets: text, time of creation, tweet id, and language.

The Jozef Stefan Institute is organizing this activity and is responsible for the collected and processed data and its accessibility to the project partners. The data collected is stored at the Jozef Stefan Institute for the time of the analysis processes, which is at most up to the ending date of the RESILOC project. The data is not transferred to any third parties (meaning no one from outside the RESILOC consortium has access to the data).

Data subjects can contact the regarding personal data inquiries at luka.virag@ijs.si . For further information on this activity and on the RESILOC project data subjects can contact Martin Žnidaršič at martin.znidarsic@ijs.si

Jozef Stefan Institute Data Protection Officer: luka.virag@ijs.si

Jozef Stefan Institute Contact Person for Social Media Analysis: martin.znidarsic@ijs.si

Local Contact Person: Paola Lorenzoni at lrt.resiloc@comune.gorizia.it

RESILOC Data Protection Officer: datenschutz@zv.fraunhofer.de