RESILOC Information Sheet – *'Social Media Analysis to be used in Gorizia Trial'*	
Research Project Title	Resilient Europe & Societies by Innovating Local Communities
Project Background	The 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.
Project Website	https://www.resilocproject.eu/
Invitation	This information sheet presents an overview of the social media analysis activity based on aggregated information from tweets that will be part of the Gorizia Field Trial.
Purpose of the research	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.
Purpose of the specific research activity	The social media analysis activity based on aggregated information from tweets is an experimental activity with Twitter data that aims to capture specific information, such as volume of tweets and their sentiment (positive, negative or neutral) on topics that are predefined by the Community of Gorizia in its role of RESILOC community. The results are going to be used by the Local Resilience Team of Gorizia as main users of the RESILOC tools to 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 social media is potentially useful as a proxy in the resilience assessment of their community. This is a research purpose in itself. The tweets that are used are public and the 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 the social media platform and the consortium.
Research activity procedures	The data (tweets) are accessed by using Twitter API for Academic Research. For the case of the Gorizia Trials, the general query used is "Gorizia"; the query can be updated at a later stage to focus on "Nova Gorica". The tweets are then processed in the following way: 1. Textual contents of tweets are extracted to be used in the project. 2. Then, tweets are filtered according to indicated 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). Additionally, content (text) of the tweets is used for automated detection of topics. Yearly and monthly aggregates of tweets' 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. The website hosting server stores only the aggregated data that is to be provided to you as trial participants. The analysis 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 project ending date. 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 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 were 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).
Data used	The tweets that are used represent public data published on the Twitter platform (meaning they come from a publicly accessible source).
Risks	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 represents aggregates of tweets along a sentiment spectrum (positive, neutral, negative), volume of tweets and frequent words in monthly and yearly collections of tweets; it does not include individual tweets. In this sense, 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 under a research phase. This research exercise will support to 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 skew the interpretation of the results by the Local Resilience Team, hence, it may deviate the assessment of resilience in a direction that diverges from reality. However, RESILOC researchers are aware of this risk and the research team will instruct the Local Resilience Team to be mindful of it as well.
Benefits for the tweets' owners	There are no direct benefits for the tweets' owners, although it is hoped that the activity will be of public benefit by contributing to a better understanding of resilience to natural (and other) hazards and help to improve the resilience of local communities.
Privacy and Confidentiality	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 crisis and disasters. Only data that is necessary for the development of the project will be collected. In some cases, it might be necessary to collect personal information such as name, an identification number or location (Article 4(1) GDPR). In some other cases "special categories of personal data" also known as SCOPD (Article 9(1) GDPR) could be collected throughout the activities of the project. However, the processing of all personal data that are part of the tweets is limited to the procedures above and no data are stored with the purpose of back-tracing individuals.
	Data practices in RESILOC follow the principles of data minimization and use anonymization and pseudonymization techniques. In addition, data practices in RESILOC have been designed following legal requirements, as well as ethical considerations approved by ethic boards. 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, 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 Jozef Stefan Institute regarding personal data inquiries at luka.virag@ijs.si. For further information on this activity and on the RESILOC project you can contact Martin Žnidaršič at martin.znidarsic@ijs.si.
RESILOC Contact Persons	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