This document is about the algorithm used to check for the existence of identical events in the converted datasets EyesOnRussia and Civilian Harm. The following is a pseudocode explaining the algorithm steps and parameters.

**Set S** is the collection of pairs of events classified as identical, those events passed by the condition of the algorithm. They are included as accepted pairs in the supplementary materials.

**Set T** is the collection of pairs of events classified as unsure, those events are similar to some extent but the algorithm cannot determine if they are actually identical. See the supplementary materials.

**Dataset1** consists of events from EOR

**Dataset2** consists of events from CH

**FOR** event X in dataset1 and event Y from dataset2 that have identical city and date:

Calculate the distance between the coordinates of the event

Calculate the similarity of the description string

**IF** two events are backed by the same social media link & description >55% & distance <2km :

Add to S

**ELSE**:

**IF** "area" in one of the two descriptions:

**IF** similarity > 75% and distance < 2km:

Add to S

**ELSE:**

Add to T

**IF** anykeyword is in one of the two descriptions:  
 **IF** description > 55 % and distance < 1km:

Add to S

**Else:**

Add to T

As for the algorithm, we take these keywords: bridge, college, market, theater, residential, zoo, church, supermarket, university, playground, center, hotel, university, school, hospital, building, complex, house, clinic, museum, block, flat, station, factory.

The algorithm is designed to spot identical events that are found in both datasets. Doing so enables us to merge the resulting converted dataset so that we minimize duplicates in our final integrated dataset. As duplicates will lead to less accurate query results, leaving doubts about the validity of the concluded results over the resilience outcomes.

We use the following as an example to explain the details of our algorithm. The two following events were found. An event with the description "Shopping center destroyed by explosion" was found in EyesOnRussia and an event with the description "Shopping center destroyed by explosion" was found in Civilian Harm. Both events are about the same incident that happened on 7th March 2022 in Kherson with coordinates (46.672988, 32.643975) in EyesOnRussia and coordinates (46.672852, 32.643839) in Civilian Harm. Based on their coordinates, we found that these two reported events are only 183.5 meters away from each other. Without the use of the algorithm, these events will be represented twice in our integrated dataset, in the case of counting the number of shopping centers destroyed, we will falsely count this shopping center twice, where in fact it was reported to be damaged just once. Thus, events like this are stored in set S and are considered identical. These pairs are taken for the next step for integration.