Research on hate speech going viral
-data collection methodology-

1. Research aims

- identify the criteria that should be taken into consideration when mapping the types of situations that tend to generate more hate speech online
- quantify the predominance (occurrence factor) that each identified criterion (or complex of criteria) has on the probability that a certain situation would lead to hate speech going viral
- try to identify early warning signals with regards to events or situations that tend to generate more hate speech online

2. Acquiring the corpus for the research

National press will be scanned daily in order to identify **current** events that could be included in our research. In addition to this, a google news search will be performed daily, restricted to news form the partner's country. Keywords indicating groups vulnerable to hate speech will be used (e.g. Muslim, gay, refugee etc.). News published over month ago from the date of the search will be discarded.

We shall also collect data on the occurrence of the events on social networks. If partners have access to a scrapping service, they are encouraged to use it. If not, the search functionalities of the dominant social networks in each of our countries can be used.

3. Coding process

Each of the events included in the corpus of the research will be coded in the coding grill annexed to this methodology. Each event will be reserved a separate line in the coding grill. The coding grill is composed of the following items:

1. Description of the event (max 200 characters) – a brief description of the event, in English, limited to 200 characters. Information should be given as to what happened, where it happened, who was involved

2. Date of occurrence – either a single date or a timeframe in which the events took place

3. Localization of the event – a dropdown menu. Partners must select where the event took place (local level, regional level, national level, European level, other). A example of a local level event

would be a rape committed in a county in Romania; a European level event would be the refugee crisis of 2016.

4. Perceived threat (dominant) – a dropdown menu consisting of: physical harm, property destruction, theft / minor criminal behavior, cultural changes / way of life. This section will be filled with the dominant threat that is expressed in the media articles / reader's opinions regarding the event.

5. Perceived threat (secondary) – same as the above, but the second most common threat should be listed.

6. Materialized threat – a dropdown menu consisting of: verbal violence, psychological violence, physical violence, murder, terrorist attack, rape, theft, other antisocial behavior. This section will be filled out only if the event had actual implications in real life on a general population / people that can be directly harmed by it.

7. Number of victims – a dropdown menu consisting of an approximation of the number of victims. Partners should use the information offered in the media to assess the number of victims directly affected by the event.

8 Victim type (1) – a dropdown menu consisting of: women, children, men, multiple. This is to assess a first level of emotional engagement based on sensibilities relating to traditional family roles.

9. Victims type (2) – a dropdown menu consisting of military / law enforcement, politician, public official, highly esteemed professions / influencer / regular citizen. Same as above, but correlated to sensibilities related to social roles of professions. By influencer, we mean any public figure that has the capacity to emotionally or rationally influence people (actors, pop icons, vloggers etc.). Highly esteemed professions will be assed based on the national context of the partners, but they tend to include teachers, doctors, professors, nurses etc.

10. Perceived characteristic of the perpetrator – a dropdown menu consisting of: ethnicity, religion, gender, sexual orientation, economic status, race. The main trait that describes the (alleged) perpetrator's perceived "otherness" compared to the general population / victims. The perpetrator must not necessarily have this trait, but the general public must think that the perpetrator does (e.g. a person might not be of a certain ethnicity, but the readers assume that she or he is).

11. Perpetrator's relation with country of attack – a dropdown menu consisting of: citizen, asylum seeker, immigrant, descendent of immigrants (up to 3rd generation), non-citizen. This is used to asses the (alleged) perpetrator's perceived "otherness" in relation to the state in which the event took place.

12. Minority status of the perpetrator – a dropdown menu consisting of N/A, majority, historical minority, new minority. A more in-depth look at the (alleged) perpetrator's perceived otherness. If the (alleged) perpetrator is an asylum seeker or an non-citizen of the country in which the event took place, N/A shall be filled in.

13. National context - a brief description of no more than 500 characters of the national context in which the event took place should be offered. It should cover similar events that have happened close to the date of the one being analysed, or that have contributed to the social climate in which the event took place, general stereotypes regarding the characteristics of the (alleged) perpetrator, etc.

14. Global context – same as above, but at a global scale.

15. Date of first identified articles

16. Number of press occurrences – a dropdown menu consisting of 1-20, 20-100, more than 100. An estimate of the number of press stories generated by the event will be offered.

17. Duration of media interest – a dropdown menu consisting of less that a week, 1-2 weeks, 2-4 weeks, more than a month. An estimate of the period of time in which the media has focused on the event and has produced content related to it.

18. Channel of communication (dominant) – a dropdown menu consisting of TV, written press, social networks. Partners will choose the most important medium through which the event was covered. Importance is expressed in terms of the number of audiences reached and the perceived impact on the audience.

19. Channel of communication (secondary) – similar to the above

20. Influencers involved in generating negative emotion (dominant) – a dropdown menu consisting of: mainstream politicians, extremist politicians, extremist influencers, media / entertainment public figures, social network influencers. If negative messages were spread, partners will select the most important category of people that have done so. Importance is expressed in terms of the number of audiences reached and the perceived impact on the audience.

21. Influencers involved in generating negative emotion (secondary) – similar to the above.

22. Dominant type of articles – dropdown menu consisting of informative, opinion, satire. Partners will evaluate what type of articles dominated the media coverage around the event.

23. Dominant tone of articles – dropdown menu consisting of very negative, negative, neutral, positive, very positive. Partners will evaluate what was the tone that the majority of articles had towards the group that is vulnerable to hate speech, not towards the perpetuator(s) involved in the event.

24. Articles with hateful comments on newspaper forums – dropdown menu consisting of: minority, majority, almost all. This is to assess the predominance of articles to which at least one user has generated comment that can be considered hate speech. Partners are to select minority if less than half of the articles have had hate speech comments and almost all if over 90 % of the articles have

had hate speech comments. If comments have been deleted, partners can try to evaluate if the deleted comment contained hate speech based on the other comments surrounding it.

25. Rate of hateful comments on affected articles – a dropdown menu consisting of: almost non-existent, minority, about half, majority, almost all. Partners must rely on their common sense knowledge and assess the median prevalence of user generated hate speech comments on the articles that contain at least one such comment.

26. Social network posts with hateful comments – similar to 22. Articles with hateful comments on newspaper forums, but applied to the relevant social networks in each of the partner's countries

27. Rate of hateful comments on social network posts – similar to 23. Rate of hateful comments on affected articles, but applied to the relevant social networks in each of the partner's countries

28. Hateful user generate posts on social networks – a dropdown menu consisting of: minority, majority, almost all. Partners will assess the prevalence of user generated posts on social networks than contain discriminatory remarks / hate speech and that are associated to the event that is being coded.

29. Articles and posts with CAN messages – dropdown menu consisting of: minority, majority, almost all. Partners will assess how many articles contain counter and alternative narratives (CAN) in their comments section.

30. Rate of CAN messages in articles and posts where CAN is present – a dropdown menu consisting of: almost non-existent, minority, about half, majority, almost all. Based on the articles identified at point 29, partners will assess the rate at which hate comments have been countered or alternative messages have been deployed.