ROSiE

Training Materials for Responsible Open Science

Case study 1
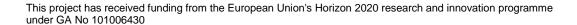### Open sharing of sensitive qualitative data

The case study is provided by Kadri Simm, University of Tartu
The Russian-Ukrainian war has caused a refugee crisis in Europe with millions of Ukrainians escaping the war zones and settling, at least temporarily, in various countries. In Estonia, scholars working with life stories want to collect narratives from recent Ukrainian refugees. Aside from the academic value of this material, the project can also be seen as a way of recording and safekeeping the tragic experiences for Ukrainians, as well as broader public. The researchers are planning to share the narratives as open data.

Questions for discussion:

1. What are the ethical issues associated with this kind of research project and what are the challenges of doing this research within the framework of Open Science?
2. What are the criteria under which such life-stories may be published as open data?
3. What information should be included in the informed consent form, especially regarding sharing narratives as open data?
4. What are risks of potential misuse of these stories? What is a risk for the refugees and their family members? What are the ways to minimize this risk?

ROSiE

Case study 2

Open sharing and reuse of sensitive data from online sources

"The Gamergate controversy erupted during the summer of 2014 and quickly engulfed the online gaming community in an intense debate about what and who belonged in gaming culture. The conflict quickly escalated into another battlefront in the culture wars involving gender, identity, and political beliefs (Hathaway 2014; Wagner 2014). Overt harassment of women and others who were critical of aspects of gaming culture quickly became a dominant feature of online forums, Twitter, and other media in which discussion about gaming culture occurred. Rape and death threats were sent to prominent game developers and journalists, some of whom were also doxxed (had personally identifiable information such as home addresses disseminated online). This harassment and other extreme expressions of hatred and intolerance quickly eclipsed the alleged ethical issues in gaming culture that supposedly was the impetus for Gamergate."

Research done by Suomela et al. included building a primary data archive comprising data collected online and including numerous instances of hate speech. They describe the ethical challenges as following:

"Privacy and reputation were the two biggest harms to research subjects that we discussed and attempted to mitigate in the Gamergate project. The privacy of our subjects was protected in two overlapping ways. First, the results of the research were reported only in aggregate forms, and second the sources for direct quotes, were not identified [..]. Neither method of privacy protection can completely guarantee that people will not be identified because the activity we collected and analyzed occurred in online forums like Twitter that anyone can search. A determined person could still recover the original source of a quote by searching Twitter or the internet, so the results could not be completely anonymized. The question of privacy is highly fraught when it comes to research about topics such as Gamergate, which depend on the internet as the primary medium for communication. Any quote from a publicly accessible web site could potentially be re-identified after a research study has been completed."

Questions for discussion:

1. The posts gathered from social media are publicly available. If so, does it mean, that there are no restrictions for how this data can be gathered, shared and reused for research purposes?
2. Might there be any need to get consent from the social media account owners to use the data for research? Are they research subjects? On what conditions the data may be used?
3. How should the data be anonymized or pseudonymised? Are the methods suggested by the authors sufficient to protect privacy of research subjects?