

Case study 1

Open sharing of sensitive qualitative data

AUTHOR: Kadri Simm, University of Tartu

The Russia-Ukraine war that began in 2022 caused a refugee crisis in Europe, with millions of Ukrainians escaping war zones and settling, at least temporarily, in various countries. In Estonia, scholars working on life stories wish to collect narratives from recent Ukrainian refugees and publish them on an open access project platform. Aside from the academic value of this material, the project can also be seen as a way of recording and safekeeping the Ukrainians' tragic experiences for both them and a wider public. Some participants would like to openly publish their stories of war and escape even under their own names so that their experience can inform the public about what happened and allow scientists to analyse their experiences. However, while the war in Ukraine continues, there is also a simultaneous information war raging, and the refugees' relatives may still be fighting in the former. Further, some refugees might have witnessed war crimes and the possibility exists that these stories could later be used as evidence in a court of law. Further, among refugees, there are children and adolescents whose stories form part of those told by their family members.

Questions for discussion:

- 1) What are the ethical issues raised by 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 the risks of potential misuse of these stories? What is the risk for the refugees and their family members? What are the ways to minimize this risk?

Case study 2

Open sharing and reuse of sensitive data from online sources

Suomela, T., Chee, F., Berendt, B., & Rockwell, G. (2019). Applying an ethics of care to internet research: Gamergate and digital humanities. *Digital Studies/Le Champ Numérique*, 9(1). <https://doi.org/10.16995/dscn.302>

"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 battlefield in the culture wars involving gender, identity, and political beliefs [...]. 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 doxed (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 on the Gamergate controversy and including numerous instances of hate speech. They describe the ethical challenges as follows:

"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 depends on the internet as the primary medium for communication. Any quote from a publicly accessible website 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 on 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?