The Good, The Bad, and The Grey: Qualitative Management Research and Ethical Data Collection From Social Media Communities

Lakshmi Balachandran Nair
Luiss University

TEACHING CASES 2022

LUISS

Introduction.
The Context, Key Players, Scene

This case study examines data collection and related ethical considerations in the context of social media communities.

We are living in a digitalized world. The different domains of social life are increasingly getting reorganized around digital communication and social media infrastructures (Brennen & Kreiss, 2016). Especially since the commencement of COVID-19 pandemic, the use of social media and other online platforms has increased exponentially across the world (Wheeler, 2020). Social media can be defined as online spaces in which social interactions amongst people are enabled (Leppänen et al., 2013). People using social media connect and communicate with others daily through collaborative generation, exchange, and continuous modification of content (Kaplan and Haenlein, 2010; McKenna et al., 2017). At this point, it is important to note that these social media platforms are 'social' in different ways (Baym, 2011). To give an example, Facebook users mostly connect with people in their friend circles. This online platform affords multimodal data in the form of texts, photos, and videos shared by users for the perusal of other users who are familiar and friendly with them (Mikhaeil and Baskerville, 2019). Reddit, on the other hand, is a social news aggregation and discussion website. Reddit data is mostly textual. Although ascertaining strict boundaries between different types of social media platforms is difficult due to their fuzzy nature, in general Facebook is a site in which people connect with personal acquaintances (by adding them as "Facebook friends") and like-minded individuals (by joining Facebook groups of shared interests).

Prior research has classified such social media platforms into five types – blogs (e.g.: My Plastic Free Life), social networking sites (e.g.: Facebook), collaborative projects (e.g.: Wikipedia), content communities (e.g.: YouTube), and virtual worlds (e.g.: Second Life). Figure 1 provides the general definitions of each classification. These classifications, however, are not completely exclusive. Some social media platforms might fit into more than one classification. For example, Yahoo! Answers is a social networking site as well as a collaborative project (Agichtein et al., 2018; Nair, 2020). The omnipresence of these social media platforms in our everyday life has created several new opportunities for qualitative management and organizational researchers (Blank, 2017). The large quantity of easily searchable, low-cost digital data available online offers unique opportunities to researchers who explore the activities and interactions of people (Hewson, 2013).
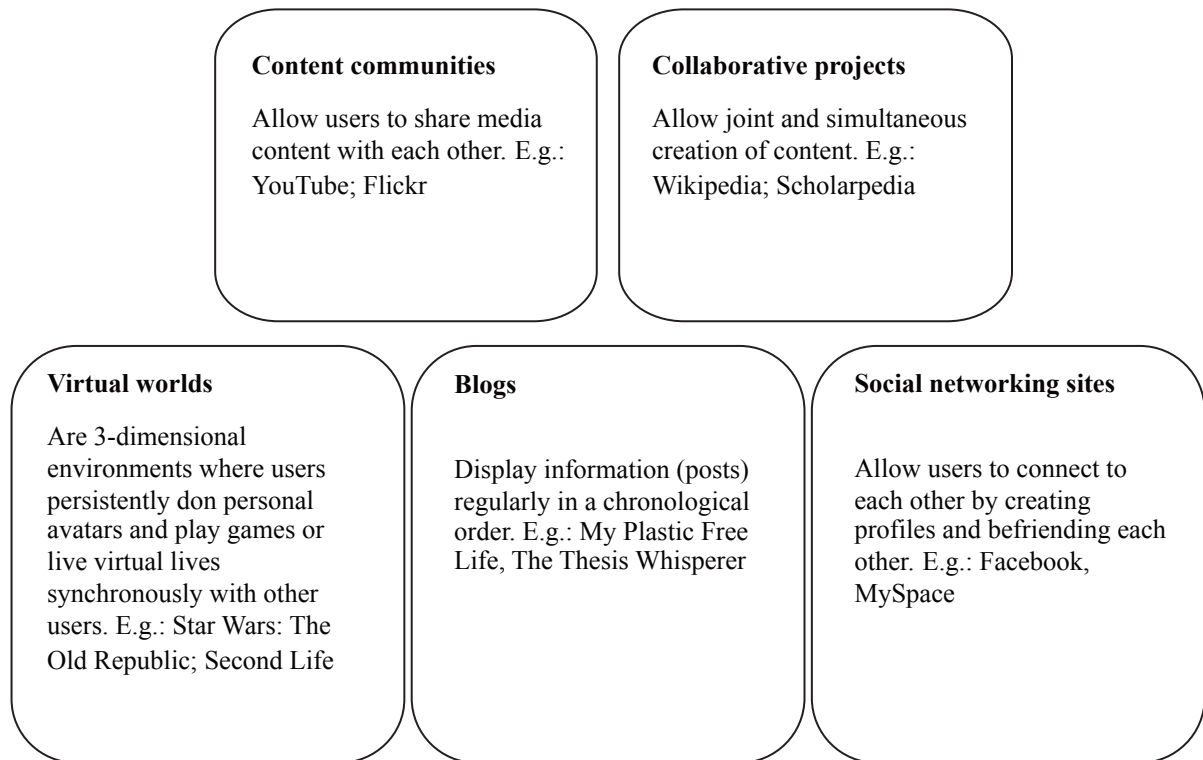
Content communities

Allow users to share media content with each other. E.g.: YouTube; Flickr

Collaborative projects

Allow joint and simultaneous creation of content. E.g.: Wikipedia; Scholarpedia

Virtual worlds

Are 3-dimensional environments where users persistently don personal avatars and play games or live virtual lives synchronously with other users. E.g.: Star Wars: The Old Republic; Second Life

Blogs

Display information (posts) regularly in a chronological order. E.g.: My Plastic Free Life, The Thesis Whisperer

Social networking sites

Allow users to connect to each other by creating profiles and befriending each other. E.g.: Facebook, MySpace

Figure 1. Kaplan and Haenlein's classification of social media platforms
Notes: Adapted from Kaplan and Haenlein (2010)

Of the aforementioned five classifications, social networking sites and content communities are the most common data collection platforms used by qualitative researchers. For instance, the social networking site Facebook provides possibilities for the researchers to come in close interaction with respondents through friendships and group memberships. In this manner, the researchers are able to understand the real-life situations of the study participants. In the context of qualitative management and organizational research, this opportunity to understand the experiences of participants aids in the production of practically relevant knowledge (Alasuutari, 2010; Leonard-Barton, 1990). In this case study, we discuss one case of data collection through Facebook. We discuss the data collection endeavors of Mila Myers[1], a master's student interested in exploring the impact of COVID-19 on the work-life balance of global female workforce. Specifically, she is interested in collecting qualitative data from FEAP (Female Employees Amidst Pandemic)[2], a Facebook community which connects working women globally during COVID-19.

1. Imaginary name.
2. Imaginary Facebook community.

What Is FEAP and Why Is Mila Interested In It?
Description

During the pandemic, many working women have lost their jobs or encountered dramatic decreases in their work hours. Many women are likewise required to juggle their personal and professional lives during the pandemic. This has proven to be a difficult task, particularly for working mothers. Due to the closure of schools and daycares, several women have had to take care of their children and educate them at home. This has resulted in many physiological (e.g.: extreme fatigue) and psychological (e.g.: stress, intense emotional labor) health issues for working women. Furthermore, the juggling of personal and professional lives in this manner also affects women's work performances. This is evidenced by the fact that the productivity of working women across multiple employment sectors has decreased since the advent of the pandemic. It is in this situation that FEAP came into existence.

FEAP was formed in March 2020, to provide a safe space for women facing problems in their work context or in juggling personal and professional lives during the pandemic. As of September 2021, FEAP has more than 7500 members across the globe. FEAP provides a platform in which women can connect with and support each other. FEAP members share concerns regarding their work-life balance and give each other tips to improve their life situations. Through chats, posts, and virtual meetups, FEAP members share relevant information and job opportunities with each other. FEAP thus makes its members feel as part of a group in which their concerns are heard, acknowledged, and discussed. Although FEAP is visible to the public, only the members are able to see who is part of the group and what is posted in it. Two group administrators assess membership requests and grant access to individuals who are interested in joining FEAP.

It is in this context that Mila comes across FEAP. She found the group accidentally while browsing through Facebook. Mila is working on a thesis exploring working women's experiences regarding work-life balance during COVID-19. She has already conducted a preliminary literature review on work-life balance and has developed a few sensitizing concepts which could suggest potential lines of in-depth inquiry into the topic. Some of these sensitizing concepts are included in Table 1. So, when she discovered FEAP, Mila promptly realized its data collection potential. Being a large social media community with members all over the world, FEAP presents a perfect platform to collect ample data for Mila's research. FEAP would provide her access to relevant Facebook posts, announcements, comments, photos, videos, links to resources, events, and occasional virtual meetups. Hence, Mila decided to join FEAP as a member.

- Time spent fulfilling personal and work-related responsibilities

- Flexible work options

- Possibility to balance work obligations and family responsibilities

- Supportive company culture, policies, procedures

- Family-friendly benefits from the company

- Work-life balance friendly financial/time-related company programs

Table 1. Sensitizing concepts on work-life balance
Adapted from Lockwood (2003), for illustrative purposes only

Data Collection Through FEAP. Description

Mila informed the FEAP group administrators that she is a working woman. She did not want to disclose her student status or her actual intentions to the administrators, since she thought that doing so might prevent her from getting access to FEAP. After joining FEAP, Mila first examined some of the posts and comments of the other group members. Several such posts discussed personal experiences of how group members handled parenting fatigue while working from home. The group members found that childcare duties during the pandemic take up a lot of time and energy. Furthermore, the group members expressed feeling guilty for thinking of childcare in this manner. The comments received included tips for handling fatigue, good wishes, commenters' own similar personal experiences, and contact details of health services.

For binding the scope of her study, Mila decided to focus on extant social media data which included the posts and comments made by working mothers. See Figure 2 for a description of extant, elicited, and enacted data. Mila collected 50 posts initially, along with the corresponding comments. Some of the posts and comments also included group members' photos of family or friends. To further engage with the group members and elicit new data, Mila also pretended to be a working mother and posted her own (fake) experiences. Her posts received numerous comments, all of which Mila collected. Mila's main interest was in collecting data covertly in the form of Facebook posts and comments. However, to understand the experiences of working mothers during the pandemic more deeply, she decided to privately chat with some of the group members who responded to her posts. During these chats, she asked the group members to describe their average working day and how it changed since the commencement of the pandemic. During the chats, many of the group members became very friendly with Mila. Some of them also shared photographs of their family members and their homes. Mila thus accessed both extant and elicited textual and visual data from the Facebook posts, comments, and chats. She also took field notes based on her observation of the interactions happening in FEAP. These notes also constituted her data.
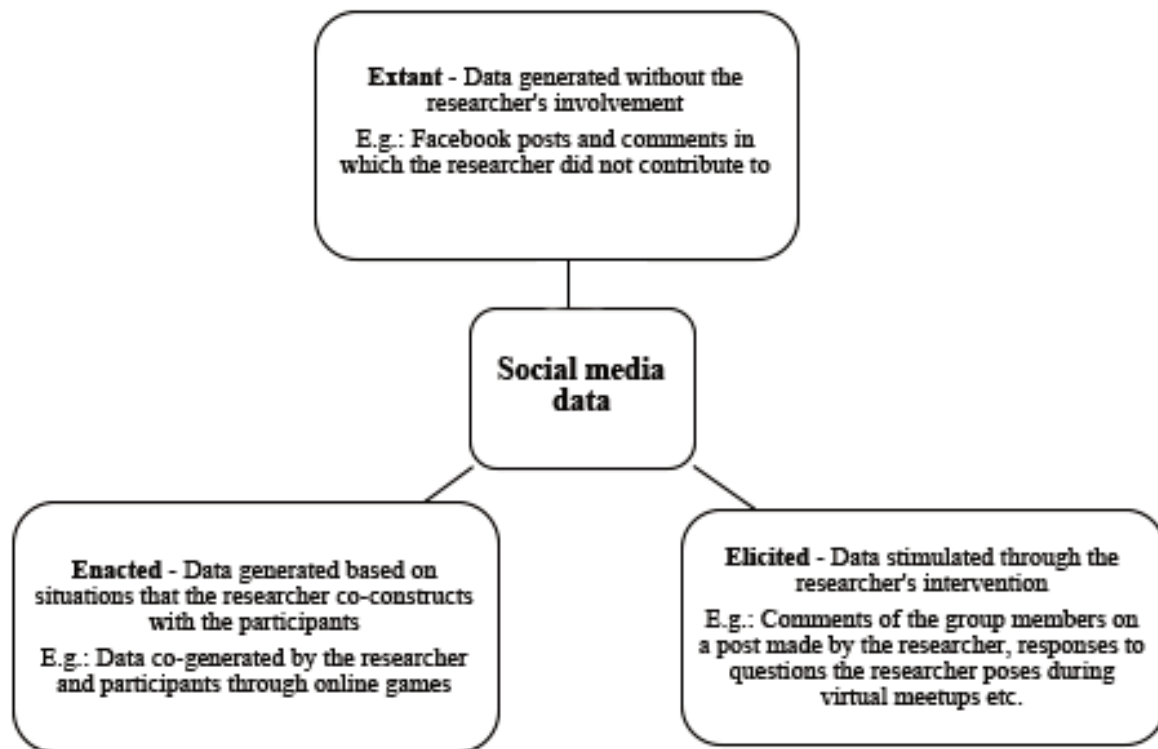
Extant - Data generated without the researcher's involvement

E.g.: Facebook posts and comments in which the researcher did not contribute to

Social media data

Enacted - Data generated based on situations that the researcher co-constructs with the participants

E.g.: Data co-generated by the researcher and participants through online games

Elicited - Data stimulated through the researcher's intervention

E.g.: Comments of the group members on a post made by the researcher, responses to questions the researcher poses during virtual meetups etc.

Figure 2. Types of data in Facebook
Note: These types of data are not exclusive to Facebook. Other social media platforms and even non-digital platforms could offer extant, elicited, and enacted data.
Extant, elicited, and enacted data can be in textual, visual, or multimodal format (e.g.: textual and visual format together)

The data collection from FEAP was conducted during October-November 2020. During this period, Mila was active on Facebook. Whenever she had time, she visited FEAP and collected data. In December 2020, Mila completed her data collection and started analysing the data. While analysing the data, she ensured that the names and identifying details of the participants were anonymized. The analysis was conducted using the grounded theory-based open-axial-selective coding process. Mila coded and categorized data into open, axial, and selective codes. In the instances where the collected data was not enough to explicate the properties and dimensions of the codes, Mila went back to FEAP and collected relevant additional data. She continued data collection and analysis until theoretical saturation was attained.

Ethical Issues Involved in Data Collection From a Facebook Community

Collecting data from an online social networking site such as Facebook is pragmatic, especially for a novice researcher like Mila. However, there are several ethical issues that underly qualitative social media data collection. Firstly, there is the issue of privacy. Facebook provides its users the option to make their profiles and groups private. Such Facebook groups have high privacy protections and restrictions. Such groups provide permissions only to selected individuals (Hennell, Limmer and Piacentini, 2019). If researchers are interested in collecting data from private Facebook groups, they are required to obtain the informed consent of the involved group members. If the researchers do not obtain consent, they will breach the perceived contextual privacy (i.e., the expected level of privacy based on the specific norms of Facebook) of the group members (Brown et al., 2016). When researchers chat with or observe group members covertly, the latter could reveal private information which they would otherwise safeguard, under the belief that there is contextual privacy.

Data collection from a Facebook group might also involve the risk of harm. When observing and collecting data from a Facebook group, the researcher might encounter instances or activities which, if publicized, are harmful to the group members. For instance, a group member might complain about her boss in an emotional and negative manner, which if exposed, will affect her relationship with her boss and in the worst case even her job prospects. Furthermore, if the specific identifiable details of the participating group members (i.e. participants) are not confidential, they could be subjected to risks of harm from perpetrators. For example, if a participant is sharing a photo of her home or children in a Facebook group and the researcher uses the photo without making it confidential, anyone who might read the final research report or have access to the data will be able to deduce where the participant lives and how her children look like. This could make the participants vulnerable to risks of harm in the form of unwanted attention or sometimes even bodily harm.

Even in the case of overt observation and data collection, the researchers are responsible for ensuring that no harm comes due to their involvement. For instance, even if the group member participating in data collection is aware of the researchers' intention and still provides them with a personal photo of herself at a friendly gathering, the researchers must ensure that the participant's identifiable information is confidential. Furthermore, they must ensure that there is no inadvertent engagement of non-participants in the study. For instance, if the participant's photo includes pictures of other, non-consenting parties who are unaware that their photo is being circulated, the researcher is required to either obtain their consent or ensure that their identifying details are confidentially handled.

While collecting data from Facebook communities such as FEAP, the researchers also should ensure that their data is of good quality. Firstly, they should ensure that the data is accessible and timely. Facebook groups provide the researchers access to a large quantity of qualitative data. It is possible that the large amount of available data might overwhelm the researcher. The researchers should sift through this data and collect only the data which is of use in their project. Collecting timely data from Facebook groups during the pandemic also offers the researchers a

possibility to examine the phenomenon of interest as it is happening. Another parameter of quality is the significance of the data. The researchers must ensure that the data they are collecting is suitable for examining the phenomenon of interest. For instance, if the focus of the study is on understanding the work-life balance experiences of working mothers during COVID-19, social media posts of non-working women or working men might not be a viable primary data source. The data collected should therefore be appropriate for developing theoretical and practical knowledge which will contribute to the understanding of the phenomenon of interest and to the resolution of the research problem in hand (Nair, 2020).

Ethical Data Collection
Decision Options

Ethically grey areas

So far, we discussed the ethical issues involved in collecting qualitative data from Facebook communities. However, it is important also to acknowledge the fact that ethical issues involved in conducting qualitative social media data collection are not always "black" (bad) or "white" (good). Rather, many ethical issues do not fall neatly into either of these two categories and might fall in a less demarcated "grey" area. These grey area issues are often uglier to address since they present dilemmas to the researcher as on how to proceed with the data collection. For instance, if Mila had informed the FEAP group members that she is not a working mother and that she is in the Facebook group to observe them and collect data, the members might have been apprehensive about violations of their perceived privacy. Even if all the members were on board with the idea, knowing that someone is observing their actions and recording them would make them behave differently from how they would do naturally. The group members might start behaving in socially acceptable manners, in an attempt to not come across as a negative person to the researcher. Mila's presence might thus act as a social stimulus which would alter the behavior of the very group that she is interested in collecting data from (Lehner-Mear, 2019). These behavior alterations would lead to Mila collecting data which is neither credible nor dependable, which in turn would negatively affect the trustworthiness of the resultant research findings.

Likewise, fully disclosing the intentions of the study to the participants might affect the access to data. Group members who might perceive the data collection as a privacy violation will resist, which could lead to conflicts between the researcher and the group members or even within group members who might possess different opinions (Convery & Cox, 2012). For instance, since Mila is collecting Facebook posts and comments, which involves the interaction of multiple group members with each other, the consent refusal of even one group member would make her comments and any subsequent responses unusable. This presents a different ethical dilemma. There might be other group members who would have liked their posts and comments to be used for research and for the subsequent betterment of the situation of working mothers. These group members' wishes will not be respected if Mila excludes their posts and comments from her study due to the consent refusal of other members. Similar conflicts can also occur if the researcher decides to obtain informed consent from non-participants who might have inadvertently gotten involved with the study due to the actions of the consenting participants. For instance, imagine that one of the participants provided Mila with a group photograph of herself with her officemates. If Mila contacts these officemates to receive their consent, there is a chance that some of them might feel grudgeful towards the participant for providing Mila their photograph without consulting with them first.

Another ethical grey area is regarding the risks of harm that the participants might face during or after qualitative social media data collection. Without question, the researchers should avoid causing harm to the participants. Harms in social science research, however, are of varying types and degrees (Hammersley & Traianou, 2012). For instance, a particular Facebook post by Mila

might irritate a group member who might then proceed to comment on it. However, the harm caused by this irritation does not damage the group member or have longer-lasting consequences. The comment posted might even provide useful data for Mila's research and eventually for the betterment of working mothers' work-life balance. Similarly, some researchers also consider potential embarrassment, boredom, wounded pride etc. of participants as less serious "harms" since these feelings would not affect the physical or psychological health and well-being of the involved parties (Feinberg, 1984). Whether to risk these potentially mild harms to the participants for rich research findings is an ethical dilemma.

Assessing the degree and types of harms involved might help the researcher in addressing such dilemmas. Let us consider another example: Qualitative researchers often include direct quotes from participants as examples in their final research reports. This practice is a way of ensuring the transparency and reliability of the research findings and the concerned study. However, in the context of social media data collection, retracing the Facebook post, comment, or photograph back to the participant who provided it is comparatively easier than in the case of traditional data. Hence, researchers should ensure that the data is always confidential. In this case, making the data confidential by using pseudonyms and by not disclosing identifiable details will ensure that no potential harm would fall to the participants. On the other hand, too much confidentiality also has its drawbacks. The richness of qualitative data might be lost if all the contextual details of the study are made confidential. The researcher should assess such situations on a case-by-case basis and find an optimal solution which protects the participants from harm and at the same time does not erode the richness of the data and the research findings.

Apart from the ethical considerations, matters such as the practicalities involved in a study could also fall into grey areas. For instance, Mila is interested in collecting data which is timely. Collecting real-time data in a timely manner is a way of ensuring the high quality of studies involving ongoing phenomenon. However, even focusing too much on real-time data can present problems. Collecting the relatively small quantity of real-time data provided by social media communities will prevent the concerned researcher from understanding the longitudinal processes or the large-scale picture of the phenomenon of interest (Nair, 2020). For example, Mila is collecting data from a single Facebook community, for a period of one month i.e., October-November 2020. The phenomenon she is interested in exploring, however, is spread over a larger period of time. Focusing on data from a single source over a small period of time without triangulating it with data from multiple sources or over different time periods could limit the trustworthiness of the study.

Conclusion

The key issue that Mila has to face in this case study is to decide how to ensure the ethical conduct of her research project. Mila has to decide how to proceed with the data collection in an ethical and yet practical manner. Should she opt for informed consent and risk having access to FEAP? Should she opt for not causing mild embarrassment to the participants at the expense of losing valuable data? Should she make the data confidential to the extent that it loses its contextual richness? How would you reach a decision? Don't forget to substantiate your choice!

References

Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008, February). Finding high-quality content in social media. In *Proceedings of the 2008 international conference on web search and data mining* (pp. 183-194).

Alasuutari, P. (2010). The rise and relevance of qualitative research. *International Journal of Social Research Methodology*, 13(2), 139-155.

Baym, N. K. (2011). Social Networks 2.0. *The handbook of Internet studies*, 2, 384.

Blank, G. (2017). The digital divide among Twitter users and its implications for social research. *Social Science Computer Review*, 35(6), 679-697.

Brennen, J. S., & Kreiss, D. (2016). Digitalization. *The international encyclopedia of communication theory and philosophy*, 1-11.

Brown, B., Weilenmann, A., McMillan, D., & Lampinen, A. (2016, May). Five provocations for ethical HCI research. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 852-863).

Convery, I., & Cox, D. (2012). A review of research ethics in internet-based research. *Practitioner Research in Higher Education*, 6(1), 50-57.

Feinberg, J. (1984). *Harm to Others* (Vol. 1). Oxford University Press on Demand.

Hammersley, M., & Traianou, A. (2012). The risk of harm. In *Ethics in Qualitative Research: Controversies and Contexts*, 57-74.

Hennell, K., Limmer, M., & Piacentini, M. (2020). Ethical dilemmas using social media in qualitative social research: A case study of online participant observation. *Sociological Research Online*, 25(3), 473-489.

Hewson, K. (2013). What size is your digital footprint?. *Phi Delta Kappan*, 94(7), 14-17.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59-68.

Lehner-Mear, R. (2020). Negotiating the ethics of Netnography: developing an ethical approach to an online study of mother perspectives. *International Journal of Social Research Methodology*, 23(2), 123-137.

Leonard-Barton, D. (1990). A dual methodology for case studies: Synergistic use of a longitudinal single site with replicated multiple sites. *Organization Science*, 1(3), 248-266.

Leppänen, S., Kytölä, S., Jousmäki, H., Peuronen, S., & Westinen, E. (2014). Entextualization and resemiotization as resources for identification in social media. In *The language of social media* (pp. 112-136). Palgrave Macmillan, London.

Lockwood, N. R. (2003). Work/life balance. *Challenges and Solutions, SHRM Research*, USA, 2-10.

McKenna, B., Myers, M. D., & Newman, M. (2017). Social media in qualitative research: Challenges and recommendations. *Information and Organization*, 27(2), 87-99.

Mikhaeil, C. A., & Baskerville, R. L. (2019). Using semiotics to analyze representational complexity in social media. *Information and Organization*, 29(4), 100271.

Nair L.B. (2020). COVID-19 and qualitative social media data collection. In *The University of the Future – Response to COVID 19* (pp. 85-92). United Kingdom: Academic Publishing.

Wheeler, T. (2020, May 5). COVID-19 has taught us the internet is critical and needs public interest oversight. Brookings. https://www.brookings.edu/blog/techtank/2020/04/29/covid-19-has-taught-us-the-internet-is-critical-and-needs-public-interest-oversight/.