During COVID-19 researchers have had to adapt their research methods to ensure safe data collection practices. Consequently, many more researchers have considered the potential of social media as a tool to facilitate their research. The ease of access to social media, and potential to generate large amounts of data easily has made this an appealing research tool. However, these methods of data collection raise unique and complex ethical issues that require careful consideration. This document has been produced in collaboration between CAHSS Research Ethics Committee and the Edinburgh Centre for Data, Culture and Society (CDCS) in response to a recognised demand for further guidance on the ethics of social media research as part of the series of Research Adaptation activities undertaken by CDCS.

The document is not intended to be exhaustive in coverage, but instead raises key issues and signposts to high quality existing resources. We have also included case study scenarios to illustrate ‘real life’ examples of navigating the implementation of social media research. These case studies are intended to be illustrative, but also provide examples that may be used for facilitating further discussion regarding the use of social media for research.

Further guidance highlighting constraints and opportunities for adapting research following COVID has been put together by CDCS. These resources focus on 1) adapting research by using digital methods when employing participatory approaches; 2) conducting archival research; and 3) outreach research in pandemic times.
There is a plethora of definitions for social media (Aichner, Grünfelder, Maurer, & Jegeni, 2021) but a recent, comprehensive and often-cited definition is:

“Social media is made up of various user-driven platforms that facilitate diffusion of compelling content, dialogue creation, and communication to a broader audience. It is essentially a digital space created by the people and for the people, and provides an environment that is conducive for interactions and networking to occur at different levels (for instance, personal, professional, business, marketing, political, and societal).” (Kapoor et al., 2018)

There is an ever evolving number of social media platforms, and Figure 1 illustrates some of the different functions and platforms (adapted from Hubspot (2021)). Users can create profiles for different social media platforms or applications that then allow the user to communicate, to consume content, and participate in creating, publishing and sharing content (Obar & Wildman, 2015). The high usage of social media provides potential to access large numbers of participants, and platforms can also provide access to specific communities (e.g., PatientsLikeMe or Mumsnet) offering an appealing environment for researchers to explore.
Multiple uses of social media for research have been discussed in the literature, and different frameworks have been produced to classify the use of social media for research purposes. For instance, Bjerglund-Andersen and Söderqvist (2012) typology divides the research uses of social media into five broad categories:

- research dissemination;
- scientific discussion and networking;
- engaging the public;
- academic teaching;
- research and data collection.

The fifth category of research and data collection is the focus of this document, more specifically this guidance is focused on making use of social media to enable research, and as a source for data collection.

As highlighted by the newly updated British Psychological Society (2021) guidance on Ethics for Internet Mediated Research, social media data collection can be broadly categorised as being reactive or non-reactive. Table 1 illustrates these two categories, with definitions and example studies, and also integrates additional perspectives (e.g., Barbosa & Milan, 2019; Kozinets, 2015).

<table>
<thead>
<tr>
<th>TYPE</th>
<th>DEFINITION</th>
<th>EXAMPLES</th>
</tr>
</thead>
</table>
| Reactive / Interactive | Participants interacting with the researcher and/or research materials on social media platforms | Using twitter to recruit participants to off-social media research activities (Wasilewski, Stinson, Webster, & Cameron, 2018)  
Interviews conducted on WhatsApp (Andries & Savadova, 2021; Gibson, 2020)  
Facebook group to co-produce intervention (Buelo, Kirk, & Jepson, 2020) |
| Non-reactive        | Data about participants that were not created as part of the research are collected unobtrusively from social media platforms | Researchers take a passive, observational approach or ‘lurker’ position (Stevens, O’Donnell, & Williams, 2015)  
Digital ethnographies (Murthy, 2012)  
Data scraping from twitter (Ahmed, Bath, & Demartini, 2017)  
Sentiment analysis of twitter posts (Abu Farha & Magdy, 2021; Liu, Zhu, & Young, 2018) |

Table 1. Categorising social media research into reactive and non-reactive approaches
Social media offers many opportunities for qualitative, quantitative and mixed-methods research, and a few notable advantages are highlighted. Firstly, by employing social media in research, large numbers of participants can be reached relatively easily and quickly, enhancing researcher-participant communication. Secondly, the accessibility of social media platforms may also enhance participation in research. For example, WhatsApp is a widely used platform that is freely available, has flexible functionality, and is easy to use on a variety of devices. Consequently, using WhatsApp may ‘open up’ research opportunities to recruit participants who may otherwise have been difficult to reach (e.g. from remote locations, entailing costly travel arrangements and time constraints on data collection).

Further, using social media data also allows for analysing trends and making associations by working with large amounts of data classified as open access.
Although social media research offers new and alternative ways to undertake research, the fundamental ethical principles remain the same. However, the unique environment created by social media requires careful consideration of how best to adhere to these principles. We encourage all researchers working with social media platforms to consider carefully the ethics of their project, and apply for ethical approval through their School REC.

The British Psychological Society (2021) guidance highlights the importance of:

- Respect for the autonomy, privacy and dignity of individuals and communities
- Scientific integrity
- Social responsibility
- Maximising benefits and minimising harms

Whilst it is beyond the scope of this document to detail all of the relevant issues, some common issues are highlighted below in relation to the four ethical principles. For further information, please see the BPS guidance, Townsend and Wallace (2016) and Taylor and Pagliari (2017).

Respect for the autonomy, privacy and dignity of individuals and communities: Public or private data?

All research participants have a right to privacy, and observation of participant behaviour for research without consent should be limited to public spaces where people may expect to be observed by strangers. Whilst this distinction between public and private is easily made in physical space, the distinction becomes blurred online and presents an ethical challenge for researchers undertaking reactive research. For example, can participants’ responses to an open Facebook group be used for research purposes? This distinction is further blurred by the fact that different people have different perceptions as to whether data are public. Whilst some argue that anything posted online is ‘fair game’, a more considered approach is encouraged. The BPS guidance suggests researchers should carefully consider the likely views of the participants, and potential harm/risks to participants of using the data to determine whether it is ethically acceptable to use it, and if valid consent is required.

In assessing potential harm to participants, researchers should be mindful of potential difficulties in protecting participants’ anonymity and preserving the privacy of their data. For example in a MSc journal club where we evaluated a paper reporting on the public’s online comments on a public health news article, we were able to identify the name of a participant based on a reported quote. Researchers need to be mindful that a quick online search of a direct quote could undermine the intention to preserve anonymity. If the risk of violation of anonymity and confidentiality is high, and this violation presents a risk then this requires careful consideration and may point towards the need to ensure informed consent for use of publically available data.
**ETHICAL CHALLENGES WITH SOCIAL MEDIA RESEARCH**

**Scientific integrity – watch out for bots!**

One issue raised in the BPS guidance in relation to scientific integrity relates to the potential of ‘bots’ completing paid online questionnaires, and therefore reducing the integrity of the data collected. Strategies such as open-ended questions, captcha and skip logic are suggested to mitigate this risk.

**Social responsibility – are you invading personal space?**

The BPS guidance highlights that social responsibility includes maintaining respect for and avoidance of disrupting social structures. Researchers who venture into online spaces may not always be welcome. As noted above, the perceptions as to whether online space is public or private may differ. For some participants, the interaction with or observation of online space may be viewed as intrusive and unwelcome. In some circumstances, undisclosed observation may be warranted because the researcher’s disclosure could harm the group identity and coherence. However, careful consideration is needed as to whether the benefits outweigh the risks, and a robust ethical case needs to be made. Stevens et al. (2015) provide a comprehensive discussion of how they made a robust ethical case for undisclosed observation in the context of collecting data from an open Facebook page to enhance understanding of managing illness.

A further challenge in digital ethnographic research is ensuring consistent data collection practices. The users’ intermittent use of social media may lead to inconsistent engagement, and high traffic platforms or public groups and pages may entail high user mobility (constantly losing some of the users while adding others), making it complex for researchers to ensure that the users are aware of an ongoing study collecting their posted data or observing their online interactions.

**Maximise benefits and minimise harm**

The lack of physical presence means that it can be more difficult to ensure efforts to minimise harm are effective in an online setting. For example, verifying that participants meet a required age inclusion criteria can be challenging. Further, it can be more difficult to read emotional responses to sensitive interview topics in an online setting, and ensure participants are fully supported. For some projects, the risks may indicate that the research is better suited to off-line methods. As noted above, online research can provide a challenge to minimising harm by protecting the anonymity or confidentiality of participants. Researchers should carefully consider if data may be traceable, and participants potentially re-identifiable. Strategies to mitigate this risk should be used, for example removing direct quotes and replacing with paraphrases.
In addition to the ethical challenges of using social media for research, there are a number of data protection and management issues that must also be considered.

For example, are the data collection processes secure? Where are the data stored, is that storage secure, and does it adhere to legal requirements? These issues can be challenging to navigate, but must be considered carefully. Researchers are encouraged to develop a data management plan to clearly outline how they will capture, use, and store the data in the short and long-term. Research Data Services provides further useful guidance on producing a data management plan.

When using social media platforms to generate new data (i.e., reactive research) then researchers are encouraged to firstly consider university supported platforms to ensure adherence to required data security and data protection requirements. For CAHSS, this includes MS Teams (preferred) and Zoom (N.B. Zoom should be avoided for sensitive topics, and at the time of writing the record and transcribe functions do not have University approval). The security and data protection of these platforms have been considered by experts and ‘approved’. (N.B. please check the most up to date guidance, as this can change)

If researchers plan to use a non-University supported platform to collect personal data then a Data Protection Impact Assessment (DPIA) is required and must be approved prior to initiating data collection. A DPIA will help ensure full compliance with GDPR requirements. There are examples available here (but these are not research specific); see also Appendix 1. It may also be helpful to get advice from Information Security. Keep a record of all of your decisions. Please note all active researchers should also complete Learn modules on i) Data Protection Training; ii) Data Protection for Research; and iii) Introduction to the Information Security Essentials (self enrol using this link).
It is important for researchers to familiarise themselves with ethical and legal implications of social media research. This document has drawn on the following resources, which researchers can also consult for guidance on how to adhere to ethical principles when conducting internet mediated research:

- The British Psychological Society: Guidance on Ethics of Internet Mediated research
- CAHSS Ethics Research Committee Considerations on research during the COVID-19 pandemic
- The University of York Guidelines for the Use of Social Media Data in Research
By putting together the two case studies presented below, ethical issues linked with employing social media in research are introduced. Potential solutions to navigating ethical decisions are included, as well as opportunities afforded by using different research methods that involve social media. The case studies are loosely based on real research studies that have been published, and these are cited in the context section of each case study. However, details have been changed and the discussion has been adapted for the purposes of developing guidance that is focused on ethics.

The two case studies presented provide examples of reactive social media research. The data produced in these studies is developed by means of direct interactions with the researcher or as prompted by the researcher, whose presence is active. These two case studies involve using the WhatsApp platform as part of research practices. Oyewole, Animasahun, and Chapman (2020) define WhatsApp as a social media platform that is freely available on different devices, and which facilitates mobile instant messaging. WhatsApp also allows for one-to-one and group sharing of voice messaging, and links to media, and various types of files. WhatsApp can also be used on computers or laptops because of its desktop version (WhatsApp Web). Alternatives to this platform, which provide similar access, flexibility in use and encryption are Telegram and Signal.

WhatsApp is not a university approved platform, and as such the use of WhatsApp for research purposes needs to be considered on a case-by-case basis. As use of WhatsApp may involve the collection of personal data, to ensure full compliance with GDPR requirements, it is essential to complete a Data Protection Impact Assessment (DPIA see link and guidance from UoE). We have included an example DPIA in relation to Case Study 1 that we hope can inform any future work (see Appendix 1).
In this study, researchers aimed to understand how young children in refugee families interacted with digital technologies in their home environments, in order to adapt to the Scottish culture. The researchers were interested in adopting an approach that would not interfere with the participants’ routines, and they were aware of the families’ familiarity with WhatsApp for daily communication. Therefore, they planned to conduct online interviews with the parents using WhatsApp.

In order to elicit information on children’s interactions with technologies the researchers used a method called ‘living journals’ (Savadova & Plowman, 2020). The living journals entailed that the parents documented via WhatsApp the children’s daily activities for a brief period of time by taking pictures of the children and making videos of their interactions. Parents could also provide textual and voice explanations to the researchers about the children’s activities on WhatsApp. Thus, the adults became proxy researchers, collecting data about their children in their own time, and in a context that was familiar and convenient for themselves and the children.

**Ethical challenges:** Particular challenges related to the use of WhatsApp for data collection and involvement of young children. WhatsApp became a data repository for developing activity diaries. The young children were the ones generating the data but the parents mediated that process. The parents solely communicated with the researchers via private messages on WhatsApp; however, the vulnerable status of the children and the generated data raised questions about safekeeping and data management.

**Solutions:** Although WhatsApp is not a University-approved tool, it has the benefit of end to end messaging encryption and its user-friendly interface. Further, it has no-cost and is widely available on different devices making it a very appealing platform for a great number of people. By opting for a University approved tool (e.g. Collaborate software) to collect data from this particular population, this may have led to exclusionary practices as the target user group may not easily or intuitively have been able to access such platforms. Ethical approval was sought for this study, and the researchers were able to address potential concerns through the ethical review process.

A data management plan was developed to carefully consider best practice in the creation, storage, and sharing of the data collected for this project. Additionally, a Data Protection Impact Assessment (DPIA) was undertaken to carefully consider the data management and security issues of the data, assess potential risks, and identify solutions. Please see Appendix 1 for example DPIA developed for this WhatsApp case study. In this case, the data collected and shared by parents included text messages, pictures, videos, audio messages as well as interview materials. On receipt of the WhatsApp message on University password-protected devices, the materials were uploaded to DataStore and deleted from individual devices. Any personal information about the participants (such as names, contact information) were stored separately from the research data. Some of the data collected included images of the children, which required careful consideration due to potential identifiability. Where parents consented for the images to be used for the purpose of reports and publications, then the images have been retained, but have been blurred where consent was not provided.

**Opportunities:** Flexibility is afforded by the data collection approach, without disrupting the participants’ daily routines, while also minimising the influence of the researcher. Proxy ethnographic approaches (Andries & Savadova, 2021; Plowman, 2017) support the remote inclusion of children in research, which can be mediated by parents and guardians who are familiar to the children.
The researchers aimed to understand the potential of digital technologies to improve approaches to civic engagement with young adults in an African context (Colom, 2021). The researchers were keen to employ an approach that would afford high ecological validity, whilst also allowing for a participatory approach and remote data collection. WhatsApp was considered to be a familiar and accessible platform for conducting research because of its presence in the daily lives of the target population, providing a practical solution to organising online focus groups. The researchers created a WhatsApp group with interested participants, for which written informed consent was obtained also via WhatsApp, but in one-to-one communication. Synchronous written conversations were prompted by the researchers in the WhatsApp group, and organised at time-periods that the participants agreed with. Asynchronous discussion was also facilitated by the platform, allowing the participants who may not be available for synchronous discussion to reflect on the other participants’ input, and provide their own.

**Potential ethical issues:** When multiple participants added their input in the same online repository, the data that they were generating as well as their contact information became visible to the other participants in the online group. The links between the participants’ identities and their views were easier to make in the online space. These would remain accessible to the participants for a period of time, as opposed to what would occur in focus groups that take place in physical places. This raised questions about the responsibilities of the participants towards their online privacy and data protection, which may not be very straightforward because of the novelty of the method.

**Solutions:** As with the previous case study, the researchers need to familiarise themselves with ethical considerations related to users’ online privacy and data protection aspects. This can be achieved by completing training on Data Protection provided by the University, seek advice from the Ethics Committee and by completing a DPIA, if a relevant one is not available.

The need for respecting the participants’ privacy and the risks associated with that need to be discussed with the participants when obtaining consent, emphasising that what will be discussed in the WhatsApp group for the purposes of the focus groups should not be shared with other individuals online. This is particularly important especially when the topic of research is sensitive, when the participants may not know each other prior to the research, as well as to avoid compromising the participants’ identity or identifiable information about them (e.g. their visible phone numbers). These principles should be reiterated by the researchers in the WhatsApp group, to remind the participants about their ethical responsibility.

**Opportunities:** Employing such an approach to adapting focus group research to remote data collection has temporal and spatial benefits, minimising the burden on participants and having the potential to increase participation. Participants would not need to make special arrangements to travel to certain physical locations, eliminating potential travel costs. Organising focus groups remotely has the added benefit of fostering asynchronous conversation, as well as making it easier for the researcher to unravel the participants’ contributions for analysis purposes.
REFERENCES


Which social networks should you focus on. https://blog.hubspot.com/marketing/which-social-networks-should-you-focus-on


References


This was part of a broader series of guidance documents created by the Centre for Data, Culture & Society and the College of Arts, Humanities and Social Sciences Research Ethics Committee for researchers in CAHSS who are pivoting to remote and digital research methods.

This document was written by Dr Valentina Andries and Dr Ailsa Niven, with support from Charlotte Smith.
CONTACT
E: CDCS@ED.AC.UK
WWW.CDCS.ED.AC.UK
@EDCDCS

CITE THIS PAPER
SOCIAL MEDIA RESEARCH: ETHICAL GUIDANCE FOR RESEARCHERS AT THE UNIVERSITY OF EDINBURGH (SEPT 2021)

REPORT, CENTRE FOR DATA, CULTURE & SOCIETY, THE UNIVERSITY OF EDINBURGH: HTTPS://EDIN.AC/3HTVAHB

IF YOU REQUIRE THIS DOCUMENT IN AN ALTERNATIVE FORMAT, SUCH AS LARGE PRINT OR A COLOURED BACKGROUND, PLEASE CONTACT LISA OTTY, CDCS@ED.AC.UK