Handbook of Ethnography in Healthcare Research

Paul M.W. Hackett, Christopher M. Hayre

Ethics of online research with human participants

Publication details
Jeff Gavin, Karen Rodham
Published online on: 03 Dec 2020

How to cite: Jeff Gavin, Karen Rodham. 03 Dec 2020, Ethics of online research with human participants from: Handbook of Ethnography in Healthcare Research Routledge
Accessed on: 17 Jan 2022

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
ETHICS OF ONLINE RESEARCH
WITH HUMAN PARTICIPANTS

Jeff Gavin and Karen Rodham

Introduction

In 2017, we wrote a short think piece on research ethics in the digital age (Gavin & Rodham, 2017). We had been invited to write that article for two reasons. First, our experience of conducting research in the online world. Between us we have worked on projects ranging from: online support for eating disorders, self-harm, and persistent pain to studies exploring dating, sexting, online intimacy, and revenge porn. We have also collected data using various online sources such as public forums; private bespoke forums; social media such as Facebook, Instagram, and Twitter; online dating sites; as well as traditional “pen and paper” surveys disseminated and completed via online platforms, such as Qualtrics. And second, when we first began to use the online world in our research, we quickly realized that ethics panels were (in our opinion) overly anxious about the ethics of doing so. They seemed to struggle to understand this kind of research, not just from a practical but also an ethical point of view. As a consequence, alongside our research, we decided that we ought to write about the ethics of conducting research on (and in) the online world.

What is online research?

Put very simply, we define online research as research which is conducted in or by the Internet or in or by digital social media. The British Psychological Society (2017, p. 3) uses the term “Internet Mediated Research”, and broadly defines it as “any research involving the remote acquisition of data from or about human participants using the internet and its associated technologies”. So, a researcher may decide to conduct an online study – perhaps running a survey using the online survey platform Qualtrics. Maybe they will prefer to run focus groups with people from all over the world joining in. Maybe they will collect data that they can subsequently analyze – for example, Instagram posts, Tweets, snapchats and so forth. Indeed, the Internet and digital social media present researchers with myriad opportunities to recruit and collect data from a diverse range of participants (e.g. Brownlow & O’Dell, 2002; Roberts, 2015; Skitka & Sargis, 2006), in ways that are often cheaper than traditional methods (e.g., no travel costs, no postage costs). The online environment also allows researchers to observe behaviour and communication (e.g. Kraut, Olson, Banaji, Bruckman, Cohen, & Couper, 2004).
What is wrong with existing research ethics guidelines?

The online world is continually evolving and has multiple layers which bring the potential for unintended consequences. This means that it is impossible for there to be a clear set of all-encompassing rules. Think for a moment about fake news (the intentional presentation of misinformation) and the more recent “deep fake” video examples. A deep fake video consists of manipulated videoclips in which someone else’s face is inserted into pre-existing videos frame by frame. As a short aside, the videos are called deep fakes after the Reddit user called “deepfake” who first created them (Guera & Delp, 2018). It is not always easy to recognize fake news and deep fakes are, as is suggested by their name, deeply hidden and almost impossible to determine as “fake”. As such, this creates a challenge for future researchers in the form of a philosophical and ethical question: “What is real?” We do not intend to answer this complex question here. Instead we have argued (Gavin & Rodham, 2017) that researchers need to accept that ethics for our digital age requires the development of a different mindset, one that maintains the central ethical mantra of “do no harm” – but does so not through traditional clear cut “if-then” rules and regulations but through the process of solving puzzles. Indeed, just as in “real life”, we cannot control for all eventualities; in the multi-layered online world we need to think about different questions: What counts as data? Whose permission do we need to seek to use the data? What data, even if we can see it, might be considered private? How do we define open access? What happens if someone who is a private user links and contributes to a multi-site discussion? Does this then render his or her contribution public? (See Box 3.1)Whose permission do you need to seek? How would you solve this ethical puzzle?

We have already mentioned the difficulties ethics panels have had when assessing our applications for ethical approval. It is also clear that our professional bodies have grappled with similar dilemmas. They have struggled to produce ethical guidelines that can keep abreast of the fast-changing online world. In 2002, the British Sociological Association (BSA) decided that it was more appropriate to put the onus on the researchers themselves, as part of their professional competence, to keep abreast of developments in online research:

Members should take special care when carrying out research via the Internet. […] Members who carry out research online should ensure that they are familiar with ongoing debates on the ethics of Internet research and might wish to consider erring

---

Box 3.1 Illustration of the multi-layered online world, example adapted from Alexander (2008)

A YouTube member uploads a video. Others comment on this video, which is subsequently discovered by other Internet users through social aggregators and search services. These people add comments to the original video entry (which they might link to from their own YouTube, Facebook or Twitter accounts via “liking” or “sharing” the video or “following” the original poster), view the video, and add further comments on YouTube, thus intensifying and contributing further to a networked discussion across multiple sites, with multiple authors and with new text, hypertext, and audio-visual content.
on the side of caution in making judgements affecting the well-being of online research participants.

(BSA, 2002)

More recently, the British Sociological Association (2017) published their “Statement of Ethical Practice” which has a separate annex entitled “Ethics Guidelines and Collated Resources for Digital Research”. Both documents can be downloaded from the Association’s “Guidelines on ethical research” website. These two documents, and particularly the annex, offer a more in-depth exploration of the possible ethical issues surrounding online research. But the conclusion is very similar: that the field is fast moving and that there are likely to be unintended, and possibly currently unimaginable, consequences of researching in and on the online world. And, as a result, the researchers themselves, need to be willing and able to engage in an ethically appropriate manner:

[…] we should remember that ‘the fields of internet research are dynamic and heterogeneous [as] reflected in the fact that as of the time of this writing, no official guidance or ‘answers’ regarding internet research ethics have been adopted at any national or international level’ (AoIR, 2012: 2). Aside from ever-changing technological contexts, and the unstable public/private distinction, the AoIR [Association of Internet Researchers] also identifies the complex and unresolved relationship between data and persons: ‘Is one’s digital information an extension of the self?’ The data/person relationship is a central issue for research ethics, as ethics aim to minimise harm, and harm is typically understood in relation to ‘persons’ (2012: 3, 6–7). This all leads back to reiterating a dynamic, situational, process-based and dialogic approach to ethical digital research; where you anticipate that unforeseen situations, issues, and technologies may arise, and you are prepared to engage in an ongoing way.

(BSA, 2017, p. 8)

Similarly, the British Psychological Society’s (BPS) second edition of the Internet-Mediated Research Guidelines (2017) recognized that as technology advances, changes, and grows, it extends the opportunities for research whilst at the same time, introduces extra complexities in ways that might not at first be obvious. The second edition BPS Internet-Mediated Research Guidelines (2017) highlight the issues facing researchers and the need for ethical guidelines not to be used as a rule book, but as a set of guiding principles. In short, it is not possible to have a set of ethical rules that can deal with all situations. How then, as researchers, do we uphold the essence of our ethical principles when conducting research in and on the online world?

In this chapter we set out to address this question by returning to the three universal principles that underpin the notion of “do no harm” with respect to how we go about conduct ethical research:

• respect for the autonomy, privacy, and dignity of individuals and communities;
• scientific integrity;
• maximizing benefits and minimizing harm

We will structure our chapter according to these universal principles and will share good and bad practices to demonstrate how we can work towards maintaining the overarching goal of ethical research: do no harm to your participants.
In the online world, working out what is public and what is private is not straightforward. How do we label different types of information that could become data if we collect it? How can we be sure that someone posting information online knows or expects it to be public? Does it matter if information is posted (and collected for research) on discussion forums, Twitter, YouTube or Facebook? Should researchers be asking themselves the following question: just because we can collect online information, should we? For example, at the time of writing, the default setting on Qualtrics (a simple-to-use, web-based survey tool) is set to collect data on the precise location of respondents. This is not unique to Qualtrics; many data harvesting software packages for collecting Tweets have the same default setting. Failure to turn off such features means that it is possible to conduct analyses which may reveal people’s personal characteristics and potentially their identity, which they may have assumed to be private.

In the UK, the recent introduction of the General Data Protection Regulations (GDPR) published by the Information Commissioner’s Office (2018) has highlighted the problem related to collecting data “just because you can”. In essence the GDPR is Europe’s new framework for data protection laws. The UK Research and Innovation organization (UKRI) has a very useful summary for how the GDPR impacts researchers. One of the requirements of the GDPR is that researchers must have a clear rationale for collecting different types of personal data. Personal data is described as information that relates to an identified or identifiable individual. This could be a person’s name or might include other identifiers such as an IP address or a cookie identifier. The key issue here is that if it is possible to identify a person directly from the information being collected and analyzed, then that information may be personal data.

In addition to ensuring that we consider carefully what counts as personal data, the GDPR also introduces the notion of data minimization. This refers to the expectation that only data which is relevant to the research in question will be collected. In other words researchers need to be able to demonstrate that they have appropriate processes in place to ensure that they only collect and hold the personal data they need. In other words, gathering or collecting information just because you can access it is not permissible unless you have clear rationale that links back to your research question. In the same way that researchers must, in effect, seek consent from their ethical bodies to collect data, so too should they seek informed consent from their participants.

Under GDPR, even data collection methods that may previously have seemed straightforward, quite rightly require further thought. For example, if we choose to collect data by means of a quantitative online survey, how can we be sure that a participant who shuts down their web browser has simply decided that they no longer wish to participate or that they have decided to withdraw consent? How can we be sure it is still okay for us to use the data we collected before they shut down their web browser? The BPS (2017) suggests that these kind of potential issues should be anticipated and withdrawal procedures made as clear and robust as possible. One example offered is to display a clearly visible “exit” or “withdraw” button on each page of a survey. If participants were to click on this button they could then be taken to a debrief page and a tick box section asking participants to confirm they are still happy for their data to be included in the study.

For qualitative research there are different considerations. If a member decides they no longer wish to participate and withdraws from the discussion in the same way those participating in
a face-to-face focus group have the option to do, what should a researcher do with the data? Omitting that person’s contributions is likely to render the other participants’ words meaningless. A focus group is, by its very nature, interactive; what one member of the group says will build on the contributions of others. Removing one voice can mean that the whole group discussion is unusable. However, as long as an information sheet is explicit about the process of withdrawal, this issue can be prevented. For example, it is common for those being invited to take part in face-to-face focus groups to be told that they can withdraw from the group discussion itself, at any time, but that all contributions they have made up to that point will be included in the analysis. The same clause could easily be included in information sheets for online focus groups.

Ethical concern about consent in the context of online studies is illustrated well by the following study where researchers had sought consent from what they considered appropriate sources. In 2008, a team of experienced researchers in the USA published a study based on data taken from the Facebook profiles of the entire cohort of a U.S. university (Lewis, Kaufman, & Christakis, 2008, see Zimmer, 2010 for review). Their focus was on how friendships and tastes develop over time. The researchers had permission from both Facebook and the university in question and had also received ethical approval from the relevant boards. They downloaded each student’s name, gender, major, and their network of friends, including who was tagged in their photos. They collected information on tastes, political views, and romantic interests, and inferred students’ race from their photos and group membership. Students from the same university were used as research assistants to access this information because, in 2006, Facebook privacy was predominantly based on university networks, and only those in the same network could see your profile. This meant that these research assistant-students had access to the profiles that the researchers themselves could not access.

In accordance with ethical guidelines, permission to use the data was not sought from the users, but the data were anonymized and all identifying information (such as names and identification numbers) removed from the published data. Unfortunately, as regular social media users know, it is not difficult to use indirect means to find somebody on Facebook. In this case the university was identifiable by its unique characteristics (e.g., the number of students, the combination of degrees offered), and from here individual students could easily be identified, particularly if they were in some way unique (such as the only female Latvian law major, to take a hypothetical example). The net result was that participants did not consent to their data being used, and worse, their anonymity was not protected.

Scientific integrity

Ensuring research maintains the principle of scientific integrity means that it “should be designed, reviewed, and conducted in a way that ensures its quality, integrity, and contribution to the development of knowledge and understanding” (British Psychology Society, 2014, p. 9). If a research study is not designed well or is conducted poorly then it is effectively a wasted opportunity. A poorly designed or implemented study will not collect high quality data. As such, it wastes resources (financial, equipment, time) and more importantly devalues the contribution of the participants and, in so doing, affords them a great lack of respect. At worst, such research may lead to inaccurate or misleading information being shared or influencing subsequent research, policy decisions, and so forth, and as such, it can have the potential to cause harm. Researchers must therefore ensure that their work meets high quality, robust scientific and scholarly standards.
**Scientific integrity vs participant safety**

The underpinning principle of scientific integrity is very clear and unambiguous. However, scientific integrity can conflict with the need to keep participants safe. For example, when conducting qualitative research the actual words used by participants are important. How people say things, the words *they* choose to use when talking about things are important. However, unlike face-to-face interviews or focus groups, collecting text from online sources brings the potential that the extracts could be placed into a search engine and the original interaction or posting identified and with that identification comes the risk that the person’s identity could be traced and revealed. For example, the BPS guidelines note that:

On a legal note, should a person find out that their online posts or traces of activity have been accessed, stored and used as research data, they are likely to have rights under the Data Protection Act to stop these data being processed if they could be linked to them personally. In many cases it is very unlikely that a person will ever find out that their online posts have been used for research purposes. However, this does not preclude the responsibility of the researcher to ensure that maximal anonymisation procedures are implemented (for example, researchers may consider paraphrasing any verbatim quotes so as to reduce the risk of these being traced to source, and participants identified). Here again, the principle of proportionality becomes pertinent: considerations of the level of risk/harm must be weighed up against scientific value, the quality and authenticity of reports of research findings, and possible practical issues too.

(BPS, 2017, p. 14)

This then provides us with a clear example of a time when scientific integrity – using the data as it was collected – may be outweighed by the need to ensure participants’ confidentiality is maintained. This might well involve paraphrasing verbatim quotes in order to reduce the risk that what someone said online can be traced back to them. The paraphrasing would be completed post analysis and so would not impact on the researcher’s ability to interpret the data they have collected; it may however, render the report of the study less convincing to the reader (and journal reviewers), perhaps even undermining the epistemological and ontological bases of the study (for example, in the case of Foucauldian discourse analysis or narrative analysis where it is language and the way in which language is used which is central to the analysis).

The decision to paraphrase (or not) also has consequences for replication. We are thinking here of the fact that there is currently a drive towards ensuring research is both open and transparent. As such, it is becoming more common for researchers to be expected to deposit their data in an open science repository. What then should we do with qualitative data collected from the online world? If we need to paraphrase it in order to protect anonymity and confidentiality of our participants (who, by the way, if the data was collected from a publicly accessible site, may be unaware that their words have been used in research), should we deposit the paraphrased text or the original text? If we do the latter, we are potentially opening up the possibility that our participants may be identified. If we do the former, the data available for other researchers to analyze is not the actual data that was collected, which defeats the purpose of open science.

A solution used by one of the authors of this chapter (JG) in disseminating his research on online dating is to illustrate his arguments with screenshots of his own (mock) online dating profile (Gavin & Griffin, 2012). Similarly, in a series of recent studies examining the perceived attractiveness of autistic males’ online dating profiles, the same author and his colleagues first analyzed the actual profiles of autistic male online daters, and then used this analysis to create a
prototypical profile to use (and manipulate) in a number of online experiments (Gavin, Rees-Evans, Duckett, & Brosnan, 2019). As it was considered ethically impossible to use profile photos of real online daters, the researchers used an “average” face comprised of a composite image created by averaging and then combining the shape and colour information of a number of individual facial photographs using specialized software.

In contrast, for our research on images of self-harm posted on a public forum (Rodham, Gavin, Lewis, St Denis, & Bandalli, 2013), creating “mock up” images would have been inappropriate. The images were the data, therefore we needed to analyze the images themselves. We took two approaches to maintaining the scientific integrity of our data whilst also ensuring participant safety. During the dissemination phase of this study, we only used images of self-harm that contained no identifying information such as faces or a distinctive background. This rendered them privately public; that is, private in the sense that they contained no identifying information, but public in that anyone could see them (Lange, 2007). Our second strategy was to publish a content analysis of the images without including the images themselves. In each of the cases cited above, ethical and methodological compromises are made in consideration with the specific research questions, type of analyses, and social context of the online site in/on which the research is conducted.

Maximizing benefits and minimizing harm

Maximizing benefit and minimizing harm refers to the aspiration that the research conducted brings the most benefit it can without harming, or, at the very least, minimizing the risk of actual or potential harm as a consequence of data collection, analysis and publication. In short, this process is about “ensuring scientific value (maximizing benefits) and taking steps to protect participants from any adverse effects arising from the research” (BSA, 2017, p. 18).

In whatever social context we as researchers work, we should be mindful and respectful of social structures. The BPS code of human research ethics (2014, p. 10) states: “unwarranted or unnecessary disruption should be avoided unless the psychologist judges that the benefits of intervention outweigh the costs of such disruption”. This holds for online research and comes back to the difficulty that sometimes arises in distinguishing between what is considered a private or a public online space by users. It is not necessarily the interventions themselves that are potentially harmful, but their possible scope for compromising the anonymity/confidentiality of participants. Researchers should consider such potential unintended consequences. For example, if researchers enter open access online spaces that are considered private by their users, their presence is likely to be unwelcome, their arrival considered invasive and potentially socially irresponsible. To mitigate this, if the proposed research is highly valued in terms of scientific integrity and potential benefits, then a researcher might feel that joining a group without disclosing that they are a researcher is an appropriate course of action. Doing so will enable them to undertake undisclosed observation and data collection whilst avoiding disrupting the space and causing potential harm perhaps to group cohesion. However, this strategy is not to be taken lightly, for it brings with it potentially serious negative consequence for those being studied. For example, Roberts (2015) shared an example of what happened when an online community learned later that they had been the focus of a research study. One community member commented:

When I joined this, I thought it would be a support group, not a fishbowl for a bunch of guinea pigs. I certainly don’t feel at this point that it is a safe environment, as a
support group is supposed to be, and I will not open myself up to be dissected by students or scientists.

(King, 1996, p. 122)

The expectations of those being researched must be considered, anticipated, and taken into account when planning research. This in itself is not straightforward:

Defining a space from the ‘outside’, based on access, and from the ‘inside’ based on participants’ experience of the social activities taking place … are two different positions that do not necessarily correspond.

(Bromseth, 2003, p. 73)

Whose perspective about access is correct? How do social media users feel about researchers lurking and gathering data? Hudson and Bruckman (2004) showed that users’ expectation of privacy often conflicts with the public setting in which their interactions take place. While they may be interacting in a public space, they behave (and it seems, expect to be treated), as if this is a private space. Conversely, Hargittai and Marwick (2016) have explored the ‘privacy paradox’ from another angle: when individuals claim to be concerned about privacy whilst their behaviour, especially online, runs counter to these concerns. From this point of view, there is a recognition that even if you personally have concerns and do all you can to maintain your privacy, once something is online it can generate a life of its own. One of their participants sums this up neatly:

On Facebook, I think it’s been drilled into me that you just have to assume anything you post is public. You can set your privacy settings at the strictest you want, but you just have to assume that anything you put out there can be made public to the world.

(Hargittai & Marwick, 2016, p. 3746)

What, then, are we as researchers to do about this paradox? As with much of this multilayered online world, there is no straightforward solution. How do we as researchers do the right thing without inadvertently doing the wrong thing? If a researcher announces their presence overtly, it is likely that how that group interacts may change. If a researcher lurks and, in effect, surreptitiously collects data, they run the risk of compromising the group if their presence is subsequently recognized.

Conclusion

At the start of this chapter we posed the question: if it is not possible to have a set of ethical rules that can deal with all situations, how then, as researchers, do we uphold the essence of our ethical principles when conducting research in and on the online world? This is an important question for us as researchers to consider, for the online world has made it both easier and harder than ever to conduct research. The landscape of the online world changes so rapidly that the British Sociological Association (2017) clearly stated that it was not possible to create guidelines that would be able to address all current and future forms of digital research that may become possible. This is an issue picked up by Kosinski, Matz, Gosling, Popov, and Stillwell (2015) who noted that when they were writing their article, the American Psychological Association’s website only listed three documents containing guidelines relating to research on the Internet, all of which had been written before Facebook came into being. This inability to keep up with the
Ethics of online research

Fast-changing online world means that the onus is on both ethics committees and researchers to approach online research with an open, curious, and pragmatic mind, that at all times has the phrase “do no harm” at its centre. In other words, the basic ethical principles underpinning research remain universal: 1) do no harm, 2) respect the autonomy, privacy, and dignity of participants, 3) maintain scientific integrity, and 4) maximize the benefits and minimize the harm of research. What is different is how these principles might be applied in a fast changing, multi-layered context with the high risk of unintended consequences. How the principles are applied and how unexpected happenings are dealt with will rely on the researchers’ and ethics committees’ ability to act carefully with due diligence with the information they have at that time. In summary, as far as researchers who conduct online research are concerned, we feel that they would do well to heed the words that have been ascribed to Maya Angelou:

Do the best you can until you know better. Then when you know better, do better.

References


