South African Journal of Information Management
ISSN: (Online) 1560-683X, (Print) 2078-1865

A framework of ethical issues to consider when conducting internet-based research

Authors:
Liezel Cilliers[1]
Kim Viljoen[1]

Affiliations:
[1]Department of Information Systems, Faculty of Management and Commerce, University of Fort Hare, East London, South Africa

Corresponding author:
Liezel Cilliers,
lcilliers@ufh.ac.za

Dates:
Received: 05 Mar. 2020
Accepted: 16 July 2020
Published: 02 Mar. 2021

How to cite this article:
Cilliers, L. & Viljoen, K., 2021, 'A framework of ethical issues to consider when conducting internet-based research', *South African Journal of Information Management* 23(1), a1215. https://doi.org/10.4102/sajim.v23i1.1215

Background: The Internet has changed the way that academia access information for teaching and research purposes. 'Internet-based research' refers to studies where data are collected from the internet, often social media sites. However, there are no definitive ethical guidelines for researchers on how to collect, analyse and use the data collected from internet-based research.

Objective: The objective of the article was to develop a framework that can be used by researchers to consider ethical issues in internet-based research.

Methods: The framework was developed from an in-depth literature review that examined recent research that investigated ethical dilemmas in internet-based research and existing guidelines that are available from universities across the globe to guide research ethics committees on how to evaluate such research.

Results: The framework that was developed consisted of five focus areas: the study population, legal issues, privacy expectations of users, data considerations and data storage. For each of these focus areas, several subareas were identified and discussed in detail to ensure that data were collected in a responsible and ethical manner.

Conclusion: The recommendation from the study is that both researchers and university research ethics committees should use the framework that was developed to guide their ethical decision-making process when collecting data from the Internet.

Keywords: internet-based research; netnography; ethics; privacy; university research ethics committees; data storage.

Introduction

There is a plethora of research focusing on the content that is available online and the devices that connect to the internet as well as the technologies that make these connections possible. However, for a lengthy period, the social interaction of humans on the Internet has been ignored (Viljoen & Cilliers 2019). To address this shortcoming, the field of netnography emerged, which provided 'new qualitative research methodology that adapts ethnographic research techniques to the study of cultures and communities emerging through electronic networks' (Kozinets 2002:62). However, Costello, McDermott and Wallace (2017) warn that netnography is often poorly understood by researchers and open to the abuse of private information of internet users.

The most contentious recent example of internet-based research gone wrong was the Cambridge Analytica scandal in 2014 where the private information of Facebook users was harvested. This information was used to build a system that could profile individual voters in the United States of America (USA) in order to push personalised political advertisements to their profiles (Hanna & Isaak 2018). The personal identifiable information of an estimated 87 million Facebook users was accessed without their knowledge or consent. Once Facebook became aware of the breach of privacy, the company did not alert the users of the breach and took limited steps to recover and secure the information posted on their platform (Cadwalladr & Graham-Harrison 2018). After the breach came to light, the debate about how technology can be used to impact society and the risks of privacy breaches for private citizens came to the forefront (Hanna & Isaak 2018). Therefore, whilst netnography has become an important research tool that enables researchers to collect data from online sources and communities to provide rich insight into online interaction, it must be done in a responsible and ethical manner (Chao 2015).

Internet-based research is governed by ethical guidelines and principles, which unfortunately have not kept up with technological innovation (Viljoen & Cilliers 2019). Kozinets (2002) does provide some ethical guidelines that are considered appropriate for (semi-)private online conversations, but not for online research that focuses on content analysis of general or passive internet sources such as social media networks, blogs or videos. The challenge for researchers is then how to apply traditional research ethics guidelines to netnography and similar types of online research. The objective of the article is to develop a framework that can be used by researchers to consider ethical issues in internet-based research.

The article discusses what internet-based research entails next and then presents the methodology used in this study. Thereafter the framework is introduced, and each subsection of the framework is then discussed in more detail. The article concludes with opportunities for future research.

Literature review and research motivation

Web 2.0 created new research opportunities, as it increased the scope, range and number of online communities that could be used for research purposes. Social media has become an important part of our lives. Data-centric applications, such as Facebook, Twitter and Instagram, are enabling researchers to collect massive amounts of data. The original purpose of these platforms was to keep in touch and share information with family and friends but has evolved since then into a business model as well. The shared messages between individuals and groups have become a rich source of data that can be used by businesses to sell their goods to a wider audience. Information propagation via online communities can influence product sales through the comments and products of peer users (Shi et al. 2019).

Online communities are comfortable sharing information about themselves on the internet (Costello et al. 2017). Netnographic data collection is typically qualitative and relies on a personal online connection between users. Costello et al. (2017) warn that netnography does pose a threat to the private information of internet users. The problem is further compounded as data collection online does not need to be confined to one internet site or even type of information. Rather, the type of data collected from online platforms will be determined by the research question. The litmus test for data collection should be data saturation when no new ideas or themes emerge (Logan 2015).

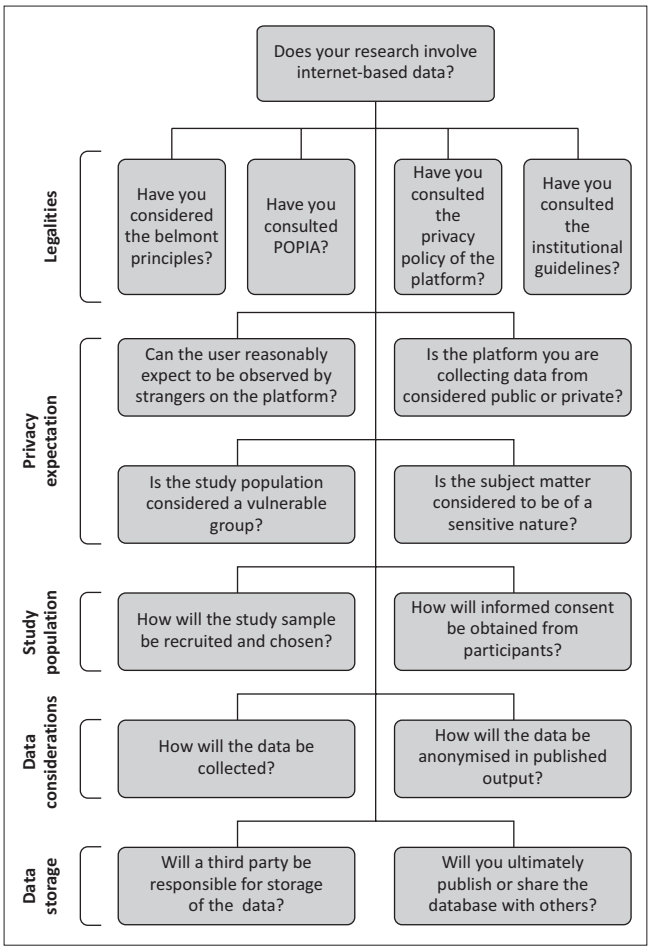
However, depending on the role that the internet may play in the research, various epistemological, logistical and ethical considerations must be considered. Researchers must define their specific internet interests and methodology in a proposal that can be considered by the relevant university research ethics committee (UREC) concerning the appropriate internet research ethical issues (Viljoen & Cilliers 2019). The first hurdle that must be overcome is the confusion among UREC members as to whether internet-based research needs to be assessed for ethical considerations, as many researchers argue that this type of research does not meet the criteria for human subject research. A human subject is defined as 'a living individual about whom an investigator obtains data through interaction with the individual or identifiable private information' (Moreno et al. 2013:708). The argument that is put forward is that internet-based research often does not require interaction with the person who posted information online as the information, whilst identifiable, is not considered private. Thus, this type of research does not require ethical screening (Viljoen & Cilliers 2019).

The UREC needs both legitimacy and authority to consider the research projects placed before them. The committee is considered legitimate, as it is the lawfully established committee with representatives from the research and wider community, which has been given as its primary task protecting potential research participants from unnecessary harm. The authority of the UREC is derived from the diverse group of experts reaching agreement through discussion and consensus on the probability of harm to the study recruits (Garrard & Dawson 2005). In South Africa, ethical guidance is the responsibility of individual URECs guided by the National Health Research Ethics Council (NHREC) (Viljoen & Cilliers 2018). The classical ethical considerations that are contemplated by UREC include issues such as confidentiality, anonymity, informed consent and dignity of the research participants. These issues remain the same for online-based research, but because of the scope, scalability and reach of the internet they become more difficult to apply (Hox & Boeije 2005). Some of the ethical issues that researchers need to address in their ethical applications to the UREC include:

- What level of privacy can users of social media sites reasonably expect when they post information on the internet?
- When can the posts be regarded as 'public information'?
- Is informed consent still necessary if the information is considered public?
- How will informed consent be obtained from the users if needed?
- How can online sampling be done in an ethical manner?

Methodology

This study made use of a qualitative, inductive approach to develop the internet-based research ethical guidelines (IBREG) framework shown in Figure 1. Constructivism is based on the philosophical perspective that knowledge is constructed rather than discovered. Making use of constructivism, new knowledge in this study was constructed by linking new information found during the literature search to prior knowledge on the subject matter (Neutzling, Pratt & Parker 2019). A literature search was used to find appropriate literature from previous studies such as articles from academic journals, books in both print and electronic format, conference proceedings and relevant university



POPIA, *Protection of Personal Information Act*.

FIGURE 1: Internet-based research ethical guidelines framework.

policies from across the world. Electronic databases such as Google Scholar, Association of Computing Machinery (ACM) Digital Library, Sage Online Journals, Science Direct, Springer Link and Sabinet Reference were used to find relevant literature. The data were collected through a literature review, which allowed the researcher to conduct several iterations through the data sources to gain a better understanding as well as identify all relevant information to the study. The inclusion criteria for the study consisted of the following: (1) research clearly defined as using the internet as the data collection tool, (2) research considering ethical issues when making use of the internet to collect data, (3) articles published between 2010 and 2018 and (4) only English literature on peer-reviewed platforms. The search strings used to search for the data included (but were not limited to) 'internet-based research', 'research ethics', 'social media research ethics' and 'netnography ethical issues'. A forward-and-backward search strategy was adopted to ensure that all useful references were included in the review. A total of 399 000 hits were found during the identification phase of the search for literature. During the screening phase, 398 430 articles were removed because of duplication, and then the remainder of the articles were screened for relevance. The remaining 570 articles were assessed for eligibility by reading the abstract, which resulted in a further 320 articles being eliminated. The remaining 250 articles were assessed by obtaining and reading the full text of the article, which left the researchers with 59 articles to include in the final literature review. The next section introduces the framework that was developed by the researcher in this regard and discusses the various parts in more detail.

The IBREG framework was developed through a literature search that identified 59 articles with direct relevance to ethical guidelines in internet-based research. From the articles were extracted different themes, issues and considerations that were used to build the framework, which consists of the key concepts of the research project and of the assumed relationships between these concepts.

Section 1: Legalities

This section will discuss the four areas identified in the legalities focus area that must be considered when conducting internet-based research. These areas include the Belmont principles, the *Protection of Personal Information Act* (POPIA), the privacy policies of the relevant social media platforms and institutional guidelines.

Belmont principles

The World Medical Association (WMA) issued the Declaration of Helsinki as a statement of ethical principles to guide those conducting health sciences researches that involve human subjects, identifiable human material or identifiable data (WMA 2001). The Singapore Statement on Research Integrity represented the first international effort to encourage the development of unified ethics policies, guidelines and codes of conduct to promote greater integrity for research worldwide (Resnik & Shamoo 2011).

Whilst both these sets of ethical guidelines were developed for medical research, they apply to other types of research as well. The National Commission created the *Belmont Report for the Protection of Human Subjects of Biomedical and Behavioural Research* in 1978. The report is a comprehensive set of ethical principles for research involving human subjects. These three core principles are identified, namely respect for persons, beneficence and justice (Department of Health 2014).

Respect for persons: The first principle, respect for persons, states that individuals should be treated as autonomous agents, and persons with diminished autonomy are entitled to protection. In order to adhere to this principle, informed consent should be sought from persons that will participate in the research. The principle further states that informed consent rests on three pillars: the extent and nature of the information given to the participant, comprehension on the part of the participant and voluntariness of participation, which is free of coercion and undue influence (Department of Health 2015; Lysaught 2004).

Beneficence: Research that involves human participants should seek to improve the human condition. The second principle refers to persons being treated ethically by

respecting their autonomy and protecting them from harm. However, beneficence also requires the researcher to maximise the possible benefits and minimise the possible harm that may come to the participants. This requires a conscious assessment of risk and benefits on the part of the researcher to determine if the risks that the participants will be exposed to are justified. The types of risks that must be considered include psychological harm, physical harm, legal harm, social harm and economic harm (Punch 2013).

Justice: Justice is the third principle and refers to the sense of 'fairness in distribution' or 'what is deserved'. The selection process of research participants must be carefully considered to determine if some classes are being systematically selected because of convenience, manipulability or vulnerability. 'Social justice' refers to the classes of subjects that should be included in a particular kind of research. This assessment is based on the ability of the members of this class to bear burdens and the appropriateness of placing further burdens on this class. Individual justice in the selection of subjects requires researchers to reasonably choose their subjects to ensure that all have an equal opportunity to participate in the research. The researcher must ensure that there are fair procedures and outcomes during the research process (Department of Health 2014; Punch 2013).

Protection of Personal Information Act

In South Africa, POPIA governs the collection, processing and sharing of personally identifiable information (PII). Although the Act was promulgated in 2013, it only came into effect in 2018 (Kandeh, Botha & Futcher 2018). There is a regulator who monitors the compliance of all individuals and organisations that collect or process personal information (Kandeh et al. 2018). However, the Act is generic in nature and does not provide enough guidance for all sectors such as mobile health and internet-based research (Katurura & Cilliers 2016).

The Act prescribes eight categories of principles that must be complied with when handling information (Kandeh et al. 2018; Katurura & Cilliers 2016). Whilst POPIA is comprehensive and protects the privacy of the individual, it does pose questions for the researcher collecting data from the Internet. The researcher is unlikely to ever meet the user face to face, which poses significant obstacles to the researcher, who must obtain informed consent, allow the user access to the data and inform the user if there was a privacy breach (Katurura & Cilliers 2016; Novation Consulting 2018; Viljoen & Cilliers 2019).

Privacy policy of the specific internet platform

The most popular internet platform is Facebook, with over 2.38 billion users (Statista 2019). Previous research has shown that Facebook users do expect their privacy on social networking sites to be maintained by the service provider (Tsay-Vogel, Shanahan & Signorielli 2018). Typically, social networks work on the principle of relationships that are granted access privileges. Networks are based on similar interests, location or even education, and the user can sign up, or 'like', the network to join. The members of the network will have access to the information of other members. Friendships require acceptance by both parties to access each other's personal data, but some sites extend the privileges to second or third degree of connection (Felt & Evans 2008).

Social networking sites make use of a terms of service (TOS) warning when a user downloads the application to a device. However, because the TOS agreement is displayed for all applications and the user cannot opt out of the agreement, it becomes meaningless to protect the privacy of the user (Felt & Evans 2008). This has necessitated an additional privacy policy that the user must agree to before they can sign up for an application. Whilst most internet-based platforms do provide the privacy policy on their website for easy reference, research has indicated that very few people will read either the TOS or privacy policy before completing their registration. The reasons for this were found to be that users view policies as a nuisance and ignored them on purpose in order to complete the actions they were busy with (Obar & Oeldorf-Hirsch 2018). Many social networking sites make it very difficult for the layperson to understand their TOS, for example Facebook's 14 000-word TOS or repeatedly changing privacy policies (Montgomery 2015).

The user does not only care about who has access to their information but also that information is disseminated appropriately (Gupta 2017). Key to this argument is the standpoint that social media users have all agreed to the TOS and privacy policy for each social media platform that they use (whether they have read it or not), and within these are clauses that explain how their data may be accessed by third parties, including researchers (Obar & Oeldorf-Hirsch 2018). However, researchers still have the responsibility to consider the extent to which the collected data may have potentially damaging effects for participants. The user's data must not be 'triangulated' in such a way that the researcher reveals potentially damaging information that the user did not originally intend to share. However, when researchers collect data on internet-based platforms, it may not be possible to notify all participants when there is a privacy breach (Viljoen & Cilliers 2018).

Institutional authority

University research ethics committees are responsible for governing the research activities of researchers in academic environments. These committees are also tasked with directing the various types and levels of ethics committees that may exist within an individual university structure. The research policy of an institution is typically based on internationally accepted standards. The most common standards mentioned across South African universities seem to be the Declaration of Helsinki, developed by the WMA, and the Singapore Statement of Research Integrity (Viljoen & Cilliers 2019).

Internet-based research, just like more traditional research methods, needs to be conducted ethically. However, at present, there is not enough guidance available to URECs on how this should be done. It is therefore imperative that research ethics committees be empowered to evaluate, monitor and ensure the ethical behaviour of internet-based research occurring within their respective universities (Viljoen & Cilliers 2018, 2019). It should also be noted that several international universities have recently developed policies for internet-based research ethics (University of California 2016; University of Cambridge 2018; University of Oxford 2016). This means that South African universities are in a position to benchmark their internet research ethics policies against international standards.

Section 2: Privacy expectation

There are four components to this layer of the framework. The first includes who can access the data of the user on the platform and if so, is their data considered public or private property? The third component deals with vulnerable groups whilst the last component considers if the data is of a sensitive data regardless of the user group. The third and fourth components will be discussed together, as the content deals with both these components.

Who can access my data?

Tsay-Vogel et al. (2018) state that social network users have an expectation of privacy when making use of social media sites. Personal information can be regarded as anything ranging from a person's name and contact details to medical and financial information. In fact, PII is any data that could potentially identify a specific individual (Els & Cilliers 2015). Mieskes (2017) conducted a literature review of privacy considerations in internet-based research and found that only 3.5% (25 of 704 publications) reported any effort to anonymise the data that were harvested.

Is my data private or public?

The individual posting on social media expects privacy despite being in a public space on a social networking platform. This contradiction creates confusion about whether it is ethically acceptable to use such data for research purposes (ESOMAR 2011). The debate regarding the expectation of privacy in online spaces has been tested on numerous occasions in the legal arena. The generally accepted principle of law is that the individual has a reasonable expectation of privacy within his or her own home but does not have a reasonable expectation of privacy in situations or acts the individual knowingly exposes to the public (Moreno et al. 2013). In the USA, federal courts have ruled that individuals do not have a reasonable expectation of privacy of information they post on their Facebook pages, as they have consented to Facebook's privacy policy. The court noted that Facebook's privacy policies plainly state that information users post may be shared with others, and that information sharing is the very nature and purpose of these social networking sites, or else they would cease to exist. However, the individual does have a measure of control by using the privacy settings of the relevant site. For example, a password-protected 'private' Facebook group can be considered private, whereas an open Twitter account through which users post their opinions using a hashtag must be considered public (Adrian et al. 2019; Moreno et al. 2013).

A general rule to remember is that the privacy policy of the specific online platform must guide the researcher as to how the data that is available on the platform may be used, whilst the profile owner is responsible for how public their data is, making use of the privacy settings. In general, tweets are generally assumed to be public, whilst Facebook posts are only public if they have been set as publicly accessible (Adrian et al. 2019). Data that are available on the internet or a social media site and can only be accessed through special permission is generally considered private. These could pertain to the private Facebook page of an individual who must 'befriend' someone before they have access to the page. Where a website requires an individual to access the data through registration, login or payment but does not restrict the data beyond these steps, data on Facebook groups, not individual pages, can be considered public as anyone with a username can access the data. The same principle applies to message boards and chat rooms (Markham & Buchanan 2002).

The researcher must also consider how he or she will recruit individuals online to participate in the research study. These efforts can therefore be compared to marketing activities. The *Consumer Protection Act* (CPA), signed into law in 2008, protects individuals from unsolicited marketing communication, as it provides them with the right to opt out of further communication and request information on where their contact details were obtained. The *Protection of Personal Information Act* also requires the individual to 'opt in' to receive marketing messages or that there be an existing relationship between the marketer or researcher and participant. The CPA also recommends that a 'do not contact registry' be kept, but the South African government has done little concerning developing this requirement (IPA 2019; Novatech Consulting 2018). Similarly, the European Union General Data Protection Regulation (GDP) requires that individuals have the right to request to be removed from a dataset. This will be difficult in the social networking arena, where online content can be copied and shared rapidly. Researchers must devise a way to handle deletion requests and check for deleted accounts longitudinally (Hunter et al. 2018).

Vulnerable groups, sensitive information and online research

There is a need to include a discussion on vulnerable groups in internet-based research, especially where the research can improve the quality of life of these groups. However, the researcher has the responsibility to protect vulnerable groups from which data may be collected. These groups include children under the age of 18, the economically

disadvantaged, racial minorities, prisoners and the elderly (Goodyear 2017; Hibbin, Samuel & Derrick 2018). These groups may not be able to give informed consent, may be easily manipulated or exploited because of their age, illness or socio-economic circumstances. Children can be considered digitally naïve and may disclose personal or sensitive information without any consideration for their own privacy or how their information will be used by others (James 2009). Research involving vulnerable groups should only be undertaken when a project cannot be carried out with a non-vulnerable group or where the research has the potential to benefit that vulnerable group (WMA 2001). The researcher is responsible for obtaining consent from the guardians of these vulnerable groups. There are several online and offline methods that can be used to obtain consent, but there are no formal guidance on how to document parental consent online (Hokke et al. 2018).

A second ethical issue that often prevents researchers from conducting research within vulnerable groups is that participants may misrepresent their age or other information to participate in the study. As the research is often conducted without any face-to-face contact, it is nearly impossible to verify personal information given by the user. Some of the suggestions provided to overcome this challenge included contacting the guardian directly to verify information, validating participant age by reviewing social media profiles, employing age verification software, limiting online research with minors to minimal risk research and conducting the research offline (Florida Atlantic University 2011; The Norwegian National Research Ethics Committees 2014).

Section 3: Study population

There are two main considerations in this section. The first is the recruitment and sampling of the study participants, whilst the second refers to how the researcher should obtain informed consent from the participants.

Recruitment and sampling

There are many opportunities for the researcher to recruit participants for research purposes on the internet. Activities such as emails, chat rooms and online advertising make it easy to reach a large audience with relatively little effort and cost (ESOMAR 2011). However, as was discussed in the previous section, one of the biggest problems is how the participants in the study population will be identified and qualified to prevent misrepresentation. This could easily happen as the researcher will have no face-to-face contact with the participant. Thus, the researcher will not be able to observe the participant reactions during the consent process (Moreno et al. 2013). This is a valid concern, but not one that is unique to internet-based research. The researcher is also unlikely to meet the respondent of a mailed survey, but the researcher must still keep in mind that traditional ethical principles must be applied and approval sought from the relevant UREC before commencing the data collection (Moreno et al. 2013).

Once the study population has been identified, the researcher must decide how he or she will sample the population. Once again, as there is no face-to-face contact, the researcher must consider the integrity of the respondents who are chosen to partake in the study. Social issues that may be considered here are the potential to duplicate respondents, respondents who maximise survey opportunities for monetary or other rewards, how vulnerable groups will be protected and the issue of representativeness of the target population. Technical issues include that respondents not be able to complete the survey more than once, how the personal data of the respondents will be protected, how the questionnaire will be designed (length and structure) and how the relevant legislation will be complied with (ESOMAR 2011; Moreno et al. 2013; University of California 2016).

In addition, individuals can also create personas, or avatars, which they may use online. Avatars are social identities that internet users establish in online communities and websites. These personas allow individuals to reveal varying levels of personal information and also allow them to navigate the virtual world as a particular character or alter-ego. These personas or avatars must be considered when sampling is conducted, as they may skew the representativeness of the study population (University of California 2016).

As is the case with traditional research, the recruitment, selection, exclusion and inclusion of participants for internet-based research should be just, fair and based on sound scientific and ethical principles. Persons should not be excluded unreasonably or unfairly on the basis of any of the prohibited grounds for discrimination: race, age, sex, sexual orientation, disability, education, religious belief, pregnancy, marital status, ethnic or social origin, conscience, belief or creed (Department of Health 2015).

Authentication measures such as personal identification numbers or other personal variables should be considered where sensitive information is collected. When research is classified as minimal risk, the standard informed consent document should confirm that the respondent is age appropriate (University of California 2016).

Informed consent

Informed consent is a key ethical issue in internet-based research. Where participants and researchers do not need to meet face to face, it is harder to establish the age and competence of the individuals and their ability to consent freely. Except where the nature of the research or participants makes this impossible, free and informed consent must be obtained from all research participants at an appropriate point in the research process (University of California 2016). There are two arguments in the literature that seem to negate informed consent in internet-based research. The first is that some scholars have suggested that informed consent is not necessary in internet-based research as it does not meet the requirements for human participant research

(Hutton & Henderson 2015; Metcalf & Crawford 2016). Moreno, Goniu, Moreno and Diekema (2013) advise that URECs must instead consider whether the study meets the criteria to qualify for human subject research. If the data collected by the study focuses on the content and not the individual who posted the content, for example YouTube videos depicting how to conduct an interview or how many Facebook pages share images of children without collecting data from the page owner, the unit of analysis is the page rather than the profile owner.

The second argument is about whether the data is considered to be in the public or private domain. If the data is in the public domain, it is freely available and thus no consent is needed (Moreno et al. 2013). The only agreement seems to be that whilst there are a number of different ways to obtain consent, there is no clear best practice for researchers to follow to seek consent from virtual subjects (Hibbin et al. 2018).

Individual UREC policies will guide the issue of informed consent. The researcher should acquaint himself or herself with these policies and abide by these rules. The following methods of informed consent could be considered by internet-based researchers, according to the University of California (2016:3):

- Internet-based questionnaires could make use of 'I agree' or 'I do not agree' buttons on the first page of the questionnaire in lieu of a signature.
- Questionnaires sent and returned via email should include a consent document that informs the respondent that submitting the completed survey indicates their consent (unsigned consent). If the consent form is printed, signed and returned to the researcher via email, this constitutes documented consent. Please note that an electronic signature may not satisfy the requirements for documented consent.
- Research that will be carried out in online communities and chat rooms should not disrupt normal group activities, but the participants and moderators should be informed that research is being conducted on the site.
- Internet-based research with minors must obtain the child's assent and the parent's permission.

Section 4: Data consideration

When data from the internet is collected, anonymisation of such data is imperative. Internet-based researchers should be well versed in the different methods of anonymising data collected on web-based platforms and should display evidence of such in their ethical applications (Hibbin et al. 2018). However, several studies have identified that large-scale aggregated data can be reidentified when the data is released (Gymrek et al. 2013; Hibbin et al. 2018; Homer et al. 2008). In addition, social media data often include photographs or videos shared via relationship in the network that is not possible to de-identify and could inadvertently reveal information about the social network or friends of the research participant (Hunter et al. 2018). Thus, it should be recognised that these measures do not guarantee privacy, and consequently every effort should be made to ensure effective protection of the stored data.

Data collection

The internet has provided a multitude of opportunities to collect data inexpensively from a diverse group of people in a short amount of time. This allows the researcher to reach groups of participants who were not previously available. It is useful to consider the difference between secondary and primary data before attempting to solve the above-mentioned issues. Primary data is collected and processed by the researcher for a specific purpose, for example to answer a research question (Creswell & Poth 2017). Secondary data was collected and processed by someone else for some other purpose but is now being used by the researcher for another reason, for example to support the findings of the current research. Research utilising secondary data that both exists and has been collected in a public, academic database, for example Google Scholar, is considered desktop research and generally does not require ethical approval from UREC (Creswell & Poth 2017). When these two definitions are considered, data that is collected on the internet or a social media site can be considered primary data, as the researchers will process the data to become information that can be used to answer the current research question. In other words, publicly available data on social media sites can be considered to be raw data (Viljoen & Cilliers 2018).

Methods of data collection

This section will consider the various methods that can be used to collect data from the internet. The first method discussed is observations.

Observations

The researcher must be sensitive to public behaviour when he or she collects data in a public space such as an online community. If the community deals with sensitive issues such as mental health, substance abuse and so on, the individuals who participate in the forum will expect privacy despite the data being available online. Participants of a forum such as a Facebook page or chatroom should be able to let the researcher know if they are not comfortable with their presence (University of California 2016). Researchers must be sensitive to this expectation and outline in their ethical considerations how they will deal with the issue of informed consent and confidentiality. A possible solution may be to create a Facebook page or chatroom specifically for research purposes and invite members to these forums. Those who choose to join the forum will receive a message informing them about the study and asking them for informed consent. This means that respondents are fully aware of the research and have consented to participate (University of California 2016).

In addition, the data that is collected cannot be used verbatim as it constitutes PII (Tsay-Vogel et al. 2018). Any details of a

person or a quote that was placed on social media that can be traced back to the person directly cannot be included in the research. To overcome this obstacle, the researcher can 'mask' the data. Masking means that raw data is transformed so that it is difficult for others to find the data online and subsequently identify the person to whom it is linked (University of California 2016).

Surveys

Researchers should outline how the confidentiality of respondents who participate in online surveys will be protected, as there are different ethical considerations from when a survey is completed in person. Normally the respondent can simply decline to answer a question, but with online surveys this may not be an option, depending on how the survey is set up. The respondents must be able to withdraw from the study or decline to answer specific questions without prejudice. Researchers should set up survey instruments in such a way that respondents can refuse to answer a specific question, for example by providing 'decline to answer' or 'not applicable' options. Where interviews will be part of the data collection method, the protocol of the method of communication should be provided to the respondent, for example voice only or video and voice (ESOMAR 2015).

Section 5: Data storage

When privacy is important to protect individual users, there must a privacy plan or protocol in place to protect the confidentiality of the users. The users can include funding agencies, human subjects, collaborators and the researcher. In small datasets, protecting PII can be achieved by removing the identifiable information and replacing it with a unique number that maps to the sensitive data in an external dataset. Hashing techniques are susceptible to a number of attacks, and all hashed data can eventually be determined. If possible, the best solution is to remove any sensitive data that is not required from the dataset prior to distribution (Hart et al. 2016).

The researcher must also plan what to do with the metadata of the project. Metadata is the contextual information required to interpret data and should be clearly defined and tightly integrated with the data. The importance of metadata for context, reusability and discovery has been written about at length in guides for data management best practices. The policy guiding metadata must include what standards and conventions will be used, how comprehensive the metadata will be and how it will be stored (Hart et al. 2016).

Where the researcher makes use of a third party to administer, store or analyse data, the potential risk of confidentiality and privacy breaches need to be considered. It is the responsibility of the researcher to make sure that the security measures to protect confidentiality and the policy used for storage by the third party are appropriate for the study (Katurura & Cilliers 2016). Respondents should be informed of these risks in the informed consent form.

Conclusion

The IBREG framework is one of the first attempts in South African literature to provide guidance on how to conduct internet-based research in an ethical manner. The problem that the framework addresses is twofold. Firstly, the researcher must navigate a complex and often confusing array of ethical issues when applying for ethical clearance for studies. Secondly, URECs are not equipped to deal with netnography research studies because of the lack of guidance. However, the interaction of humans on the internet does present new and exciting research opportunities that must be pursued in an ethical manner to produce new knowledge for this field. Future research will involve the rigorous testing of the IBREG framework in an interdisciplinary team to establish the strengths and weaknesses of the framework, whilst the inclusion of divergent perspectives would add a further layer of scope to the current framework, which can provide opportunities to expand the framework even further.

Acknowledgements

Competing interests

The authors have declared that no competing interests exist.

Authors' contributions

Both authors contributed equally to this work.

Ethical consideration

This article followed all ethical standards for carrying out research.

Funding information

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Data availability statement

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Disclaimer

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any affiliated agency of the authors.

References

Adrian, M., Moreno, M., Nicodimos, S., McCauley, E. & Vander Stoep, A., 2019, 'Research strategy for health sciences: Facebook friend request is non-differentially accepted in a diverse, young adult population', *Nursing & Health Sciences* 21(1), 71–77. https://doi.org/10.1111/nhs.12557

Cadwalladr, C. & Graham-Harrison, E., 2018, 'Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach', *The Guardian*, 17 Mar. 2018, p. 22.

Chao, B.M., 2015, 'Relevance and adoption of netnography in determining consumer behavior patterns on the web', *Scholedge International Journal of Business Policy & Governance* 2(6), 12–17.

Costello, L., McDermott, M.L. & Wallace, R., 2017, 'Netnography: Range of practices, misperceptions, and missed opportunities', *International Journal of Qualitative Methods* 16(1), 1–12. https://doi.org/10.1177/1609406917700647

Creswell, J.W. & Poth, C.N., 2017, *Qualitative inquiry and research design: Choosing among five approaches*, Sage, London.

Department of Health, 2014, 'The Belmont Report, Ethical principles and guidelines for the protection of human subjects of research', *The Journal of the American College of Dentists* 81(3), 4.

Department of Health, 2015, *Ethics in health research principles, processes and structures*, Department of Health, Pretoria.

Els, F. & Cilliers, L., 2015, 'Improving the information security in SMEs to protect customer's Personal Identifiable In-formation', in *Proceedings of the 7th International Conference on Business and Finance (ICBF)*, Cape Town, South Africa, September 09–10, 2015, pp. 75-80.

ESOMAR, 2011, *ESOMAR guideline for social media research*, viewed 15 March 2018, from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjukObg_o_gAhUCfxoKHZx_Dj0QFjAAegQIAhAC&url=https%3A%2F%2Fwww.esomar.org%2Fuploads%2Fpublic%2Fknowledge-and-standards%2Fcodes-and-guidelines%2FESOMAR-Guideline-on-Social-Media-Research.pdf&usg=AOvVaw2O2irFtwX2B3ZfA-HZzFPS.

ESOMAR, 2015, *ESOMAR/GRBN guideline for online social sampling quality*, viewed 15 March 2018, from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwj3zufV_4_gAhWE34UKHRoGBJUQFjAAegQIChAC&url=https%3A%2F%2Fwww.esomar.org%2Fuploads%2Fpublic%2Fknowledge-and-standards%2Fcodes-and-guidelines%2FESOMAR-GRBN_Online-Sample-Quality-Guideline_February-2015.pdf&usg=AOvVaw2Narp2KOsF2YbvF8nGupAK.

Felt, A. & Evans, D., 2008, 'Privacy protection for social networking platforms', in *the proceedings of the IEEE Symposium on Security and Privacy*, Oakland, CA, May 22, 2008, pp. 1–8.

Florida Atlantic University, 2011, *Guidelines for computer & Internet-based human subjects research*, Boca Raton, FL, viewed 07 July 2016, from https://www.fau.edu/research/docs/policies/research-integrity/irb_guidelines_internet_research.pdf.

Garrard, E. & Dawson, A., 2005, 'What is the role of the research ethics committee? Paternalism, inducements, and harm in research ethics', *Journal of Medical Ethics* 31(7), 419–423. https://doi.org/10.1136/jme.2004.010447

Goodyear, V.A., 2017, 'Social media, apps and wearable technologies: Navigating ethical dilemmas and procedures', *Qualitative Research in Sport, Exercise and Health* 9(3), 285–302. https://doi.org/10.1080/2159676X.2017.1303790

Gupta, S., 2017, 'Ethical issues in designing Internet-based research: Recommendations for good practice', *Journal of Research Practice* 13(2), Article D1.

Hanna, M.J. & Isaak, J., 2018, 'User data privacy: Facebook, Cambridge analytica, and privacy protection', *Computer* 51(8), 56–58. https://doi.org/10.1109/MC.2018.3191268

Hart, E.M., Barmby, P., LeBauer, D., Michonneau, F., Mount, S. & Mulrooney, P., 2016, 'Ten simple rules for digital data storage', *PLoS Computational Biology* 12(10), e1005097. https://doi.org/10.1371/journal.pcbi.1005097

Hibbin, R.A., Samuel, G. & Derrick, G.E., 2018, 'From "a fair game" to "a form of covert research": Research ethics committee members' differing notions of consent and potential risk to participants within social media research', *Journal of Empirical Research on Human Research Ethics* 13(2), 149–159. https://doi.org/10.1177/1556264617751510

Hokke, S., Hackworth, N.J., Quin, N., Bennetts, S.K., Win, H.Y., Nicholson, J.M. et al., 2018, 'Ethical issues in using the Internet to engage participants in family and child research: A scoping review', *PLoS ONE* 13(9), e0204572. https://doi.org/10.1371/journal.pone.0204572

Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J.V., Stephan, D.A., Nelson, S.F. & Craig, D.W., 2008, 'Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays', *PLoS Genetics*, 4(8), e1000167. https://doi.org/10.1371/journal.pgen.1000167

Hox, J.J. & Boeije, H.R., 2005, 'Data collection, primary versus secondary', *Encyclopaedia of Social Movement* 1(1), 593–599. https://doi.org/10.1016/B0-12-369398-5/00041-4

Hunter, R.F., Gough, A., O'Kane, N., McKeown, G., Fitzpatrick, A., Walker, T. et al., 2018, 'Ethical issues in social media research for public health', *American Journal of Public Health* 108(3), 343–350. https://doi.org/10.2105/AJPH.2017.304249

Hutton, L. & Henderson, T., 2015, "I didn't sign up for this!" Informed consent in social network research, in *Proceedings of the 9th International AAAI Conference on Web and Social Media (ICWSM)*, pp. 178–187, St Andrews, Fife, May 26 2015.

James, C., 2009, *Young people, ethics, and the new digital media*, Massachusetts MIT Press, Cambridge.

Kandeh, A.T., Botha, R.A. & Futcher, L.A., 2018, 'Enforcement of the protection of personal information (POPI) act: Perspective of data management professionals', *South African Journal of Information Management* 20(1), 1–9. https://doi.org/10.4102/sajim.v20i1.917

Katurura, M. & Cilliers, L., 2016, The extent to which the POPI Act makes provision for patient privacy in mobile personal health record systems, in *The conference proceedings of IST-Africa 2016*, Durban, South Africa, May 2016, pp. 11–13.

Kozinets, R., 2002, 'The field behind the screen: Using netnography for marketing research in online communities', *Journal of Marketing Research* 39(1), 61–72. https://doi.org/10.1509/jmkr.39.1.61.18935

Logan, A., 2015, 'Netnography: observing and interacting with celebrity in the digital world', *Celebrity Studies* 6(3), 378–381.

Lysaught, M.T., 2004, 'Respect: Or, how respect for persons became respect for autonomy', *Journal of Medicine and Philosophy* 29(6), 665–680.

Markham, A. & Buchanan, E., 2002, *Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee version 2.0*, viewed 15 March 2018, from http://www.aoir.org/reports/ethics.pdf.

Metcalf, J. & Crawford, K., 2016, 'Where are human subjects in big data research? The emerging ethics divide', *Big Data & Society* 3(1), 1–14. https://doi.org/10.1177/2053951716650211

Mieskes, M., 2017, 'A quantitative study of data in the NLP community', in *Proceedings of the first ACL workshop on ethics in natural language processing*, Valencia, Spain, April 16–17 2017.

Montgomery, K.C., 2015, 'Youth and surveillance in the Facebook era: Policy interventions and social implications', *Telecommunications Policy* 39(9), 771–786. https://doi.org/10.1016/j.telpol.2014.12.006

Moreno, M.A., Goniu, N., Moreno, P.S. & Diekema, D., 2013, 'Ethics of social media research: Common concerns and practical considerations', *Cyberpsychology, Behaviour, and Social Networking* 16(9), 708–714. https://doi.org/10.1089/cyber.2012.0334

Neutzling, M., Pratt, E. & Parker, M., 2019, 'Perceptions of learning to teach in a constructivist environment', *Physical Educator* 76(3), 756–776. https://doi.org/10.18666/TPE-2019-V76-I3-8757

Novation Consulting, 2018, *The EU general data protection regulation: Should South African organisations care*?, viewed 15 March 2018, from https://novcon.co.za/wp-content/uploads/2018/01/IPD-white-paper.pdf.

Obar, J.A. & Oeldorf-Hirsch, A., 2018, 'The biggest lie on the Internet: Ignoring the privacy policies and terms of service policies of social networking services', *Information, Communication & Society* 23(1), 128–147. https://doi.org/10.1080/1369118X.2018.1486870

Punch, K.F., 2013, *Introduction to social research: Quantitative and qualitative approaches,* London, Sage.

Resnik, D.B. & Shamoo, A.E., 2011, 'The Singapore statement on research integrity', *Accountability in Research* 18(2), 71–75. https://doi.org/10.1080/08989621.2011.557296

Shi, L.L., Liu, L., Wu, Y., Jiang, L., Panneerselvam, J. & Crole, R., 2019, 'A social sensing model for event detection and user influence discovering in social media data streams', *IEEE Transactions on Computational Social Systems* 7(1), 141–150. https://doi.org/10.1109/TCSS.2019.2938954

STATS SA, 2019, *How many active users does Facebook have?*, viewed 15 March 2018, from https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

The Norwegian National Research Ethics Committees, 2014, *Ethical guidelines for Internet research*, National Committee for Research Ethics in the Social Sciences and the Humanities, NESH, Oslo.

Tsay-Vogel, M., Shanahan, J. & Signorielli, N., 2018, 'Social media cultivating perceptions of privacy: A 5-year analysis of privacy attitudes and self-disclosure behaviors among Facebook users', *New Media & Society* 20(1), 141–161. https://doi.org/10.1177/1461444816660731

University of California, 2016, *Internet-based research*, Committee for Protection of Human Subjects, Berkeley, CA.

University of Cambridge, 2018, *University of Cambridge policy on the ethics of research involving human participants and personal data*, University of Cambridge Research Office, Cambridge.

University of Oxford, 2016, *Internet-Based Research (IBR)*, University of Oxford Research Office, Oxford.

Viljoen, K. & Cilliers, L., 2018, 'Developing awareness of ethical principles of social media research in research ethics committees', in *Conference proceedings of 6th International Conference on Ethics Education*, Stellenbosch, South Africa, October 3–5, 2018.

Viljoen, K. & Cilliers, L., 2019, *The ethics of conducting research using social media: A discussion case*, Transforming society using ICT: Contemporary discussion cases from Africa, UNISA, Pretoria.

World Medical Association, 2001, 'Ethical principles for medical research involving human subjects', *European Journal of Emergency Medicine: Official Journal of the European Society for Emergency Medicine* 8(3), 221–223. https://doi.org/10.1097/00063110-200109000-00010