



Advancing the ethical use of digital data in human research: challenges and strategies to promote ethical practice

Karin Clark¹ · Matt Duckham^{2,3} · Marilys Guillemain^{4,5} · Assunta Hunter⁶ · Jodie McVernon⁷ · Christine O'Keefe⁸ · Cathy Pitkin⁹ · Steven Praver¹⁰ · Richard Sinnott¹¹ · Deborah Warr¹² · Jenny Waycott¹³

© Springer Nature B.V. 2018

Abstract

The proliferation of digital data and internet-based research technologies is transforming the research landscape, and researchers and research ethics communities are struggling to respond to the ethical issues being raised. This paper discusses the findings from a collaborative project that explored emerging ethical issues associated with the expanding use of digital data for research. The project involved consulting with researchers from a broad range of disciplinary fields. These discussions identified five key sets of issues and informed the development of guidelines orientated to meet the needs of researchers and ethics committee members. We argue that establishing common approaches to assessing ethical risks of research involving digital data will promote consistency in the ethical standards for research, enable the smooth functioning of ethics committees, and sustain public confidence in research. We conclude with recommendations for the development of educational resources for ethics committees, data management guidelines and further public education.

Keywords Bioethics · Internet research · Digital data · Data security · Data sharing · Privacy/confidentiality · Research ethics committees · Institutional review boards

✉ Assunta Hunter
assunta.hunter@unimelb.edu.au

¹ Melbourne Law School, Grattan Street, Parkville, Melbourne, VIC 3010, Australia

² School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, VIC, Australia

³ Department of Geospatial Science, RMIT University, City Campus, 124 La Trobe St, Melbourne, VIC 3000, Australia

⁴ Faculty of Medicine Dentistry and Health Sciences, The University of Melbourne, Melbourne, Australia

⁵ Centre for Health Equity, Melbourne School of Population and Global Health, Level 2, Alan Gilbert Building, 161 Barry St, Parkville, VIC 3010, Australia

⁶ Melbourne School of Population and Global Health, 5/207 Bouverie Street, Carlton, VIC 3053, Australia

⁷ Peter Doherty Institute for Infection and Immunity, The Royal Melbourne Hospital and the Modelling and Simulation Group within the Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, 792 Elizabeth Street, Melbourne, VIC 3000, Australia

⁸ CSIRO Mathematics, Informatics and Statistics, Commonwealth Scientific and Industrial Research Organisation, GPO Box 664, Acton, ACT 2601, Australia

⁹ Social Responsibility and Ethics at CSIRO, Commonwealth Scientific and Industrial Research Organisation, GPO Box 664, Acton, ACT 2601, Australia

¹⁰ iBIONICS, F3006, 283 Boul Alexandre-Taché, Gatineau, QC J9A 1L8, Canada

¹¹ Computing and Information Systems Department, University of Melbourne, Level 03 Room 322, Doug Mcdonell Building, Parkville, Melbourne 3010, Australia

¹² McCaughey Centre, Centre for Health Equity, Melbourne School of Population and Global Health, University of Melbourne, 5/207 Bouverie Street, Carlton, Vic 3053, Australia

¹³ Department of Computing and Information Systems, University of Melbourne, Level 09 Room 15, Doug Mcdonell Building, Parkville, Melbourne, VIC 3010, Australia

Introduction: ethics and digital data

Methods for conducting human research have been transformed in recent years by digital technologies, many of which are, in and of themselves, objects of interest to researchers (Buchanan 2004, 2011; Hoser and Nitschke 2010; Kaye et al. 2015). The proliferation of digital data resources has been greatly facilitated by the use of digital devices for data capture and storage in research, and individuals' growing reliance on these technologies in everyday life (Kitchin 2014). Given widespread uptake, communication technologies involving the internet and mobile phones have become valuable tools for conducting research. They also provide scope for investigating internet-based activity (Anderson and Raine 2014; Hine 2005) and to explore community beliefs, attitudes and behaviours through a variety of social media fora. In addition, internet-based research has proved particularly useful in improving access to previously 'hard to reach' populations, such as rural communities, increasing their participation in research (See Christensen et al. 2004; Griffiths and Christensen 2007; Henderson et al. 2012).

We define digital data as information created and stored in a computer mediated environment that can potentially be transmitted as discrete information signals over the internet, and may be subsequently processed and/or stored for a range of known and unforeseen purposes. Some examples include:

- Survey data entered into computers or collected using online environments;
- Records kept by health, pathology, hospital, social and other service providers;
- Self-generated 'lifelogging' data (including metadata) emitted from mobile phones and other 'smart' appliances (e.g. Apple Watch), or generated through educational and lifestyle applications, such as fitness monitoring devices and web-based games, including data gathered from gambling, dating, and social media posts;
- Transactional and geospatial data including data generated from online records of retail purchases and the use of educational and financial services and roads and transport systems, as well as location sensing devices in public places;
- Administrative and legal data such as birth, death and marriage registrations; credit ratings, criminal convictions, and immigration and customs records.

This diversity of information sources, many of which may be linked at individual or geospatial levels, opens up new opportunities for interdisciplinary research to address complex, multidimensional social problems, and inform government policies and planning (Hargittai 2015; Holman

et al. 2008; Lane and Schur 2010; Sinnott et al. 2014, 2015; Vayena et al. 2015). The commercial sector has long recognised the value of information represented by linked data 'bundles', which have become replicable and highly portable commodities to be bought and sold (Mayer-Schönberger and Cukier 2013; Shilton 2009). The rapidity of advances in digital technologies, along with their portability and the varied stakeholders and commercial interests that may be associated with projects, are generating novel and emerging opportunities for research and prompting researchers to revisit longstanding ethical principles guiding research.

Over recent years there have been many debates concerning the ethics of using 'big data', social media posts, and other forms of digital data for academic and commercial research (Boyd and Crawford 2012; Conway 2014; Gleibs 2014; Hargittai 2015; Vayena et al. 2015; Zimmer 2018). Adapting ethical principles to address familiar and emerging risks associated with research using digital data presents complex issues and challenges. The traditional protections of Ethics Review Committees (Institutional Review Boards) and informed consent gained in a face-to-face interaction between a researcher and a research participant are challenged by research projects using big data (Zimmer 2018). This was made evident in a controversial study involving researchers from the Facebook Corporation and Cornell University (Kramer et al. 2014a, p. 8788). The Facebook researcher (Kramer) collected the data for "internal purposes" and this "may have involved practices that were not fully consistent with the principles of obtaining informed consent and allowing participants to opt out" (Kramer et al. 2014b, p. 10,779). A later editorial correction to the original article indicates that Facebook considered that it had obtained informed consent because the data collection was consistent with Facebook's Data Use Policy which users agreed to when they created an account. The Cornell University Institutional Review Board (IRB) deemed that because this was information collected for internal purposes it did not fall within Cornell's Human Research Protection Program and thus did not require that it passed their scrutiny (Jouhki et al. 2016; Kramer et al. 2014b). In effect, and without ethical review, these researchers conducted a massive psychological experiment on around 690,000 unsuspecting Facebook users when they tweaked Facebook's newsfeed algorithm with aims of exploring the effects of 'emotional contagion' by analysing users' responses to the content of posts (Boyd 2016; Selinger and Hartzog 2015). The project became public when a manuscript reporting the findings was published in an academic journal, and it ignited vociferous debate and disagreement among researchers, ethicists and legal experts. The case highlighted the ways in which the possibilities of digital data were presenting familiar ethical issues, including issues of informed consent and avoiding harm to participants, in new ways and underscored the

limitations of accepted ethical protocols. More generally, the study and its aftermath suggested that the status, reputation and integrity of academic research could be impaired if the ethical concerns that were raised were neglected or inadequately addressed (Voosen 2014).

In addition to novel research contexts and capabilities that are presented in research using digital data, the associated, ethical precepts and guidelines must anticipate increased potential for conducting international and interdisciplinary research and, therefore, must span disciplinary norms and respond to varied national legislative requirements. This paper describes a participatory project that addressed these emerging issues. The project involved interactive workshops that were designed to draw on the practical and embedded insights from researchers working across diverse settings and disciplines, from educational research to mathematical modelling and bioethics to explore key ethical issues regarding the use of digital data in research they were involved in. We aimed to identify the range of issues raised by researchers and explore key categories of issues to formulate provisional guidelines for Australian researchers. These insights and outcomes can also inform international efforts to come to grips with emerging ethical issues in the rapidly transforming fields of digital technologies and digital research. Before explaining the method for the workshops and the findings they generated, we situate our work within a growing body of research and practice seeking to understand emerging ethical risks and develop protocols to guide future research.

Responding to emerging possibilities and risks

Expanding potential to store, repurpose, harvest and ‘scrape’ data from a range of sources has been pushing researchers to reconsider whether customary strategies for meeting ethical obligations are adequate and how they can be adapted to address new risk scenarios. Researchers are familiar with core ethical principles: to demonstrate respect for participants; to obtain their informed consent to protect participants’ privacy and confidentiality; and to avoid harm occurring through participation in research. Critically, digital data can be readily stored and repurposed for future research. A key area of concern for researchers generating or repurposing digital data is to consider how these principles can be adapted to conduct secondary analyses. In situations where consent for data collection has been given, re-using or re-purposing data, in new and secondary analyses, may need to be reconsidered if the focus or scope of these analyses differ significantly from how the research was originally envisaged and explained to participants. Furthermore, many forms of digital data can be ‘emitted’ and collected without explicit consent and in the absence of any pre-specified purpose that clearly identifies

a positive value for the secondary use of this data. Researchers are increasingly able to access information in contexts where individuals may be unaware that personal information even exists in retrievable form, or that when linked with other forms of personal data, it may be possible to identify who they are. Similarly, harvesting sensitive information from public–private environments such as social media may raise ethical issues, especially for research involving vulnerable populations who may have limited understanding of the implications of disclosing personal information on these platforms (Andalibi et al. 2016; Beddows 2008; Caelear and Christensen 2010; Gustafson and Woodworth 2014; Koene et al. 2015; Lunnay et al. 2014; McKee 2013; Whitehead 2007).

Excitement at the prospect of abundant sources of information that can now be collected, stored and repurposed for new kinds of investigations should be tempered by concerns for upholding ethical obligations of minimising the risks of research. Metcalf and Crawford (2016) note that standard practices for promoting ethical research have been premised on assumptions that are less applicable to the circumstances in which digital data are generated and analysed. While there is some recognition of the need to develop ethical frameworks addressing the growing use of digital data, this point of view was not widespread among many researchers. Participants may be less likely to encounter risks of physical harm but there are increased risks of their informational privacy being breached and potential for data discrimination. Potential for risk extends to entire populations and communities of people, and increasingly abstract relationships between researchers and participants can imply diminished responsibility for the latter’s welfare (Metcalf and Crawford 2016). Further, in contexts of interdisciplinary research, it is crucial to establish shared understandings of emerging ethical risks and strategies for minimising these risks. Collaborations involving information specialists and researchers across the fields of computer studies, health and social scientists and other disciplines, can involve varying conceptualisations of relationships between individuals and data, and the ways in which ethical risks are perceived and applied (Buchanan et al. 2011; Onsrud 1995). For instance a study of researchers working with online data in the fields of computer and information sciences found consensus in identifying key ethical issues, but less agreement in conceptualising issues such what constitutes ‘public data’ and when informed consent should be obtained to use data (Vitak et al. 2016). In this study academic researchers generally demonstrated good understanding of key ethical precepts but not necessarily how these principles can be best be achieved in practice, with many researchers relying on trust in individual assessments of risks associated (Vitak et al. 2016). Growing interest in interdisciplinary collaborations underlines the importance of building shared understanding and fostering interdisciplinary communication among researchers who

otherwise are likely to be drawing on a range of assumptions, metaphors and ethical codes when working together (Ess and AoIR Ethics Working Committee 2002),¹ including researchers working outside academic settings.

We were alerted to the potential for researchers to hold diverse views on what constituted ethical approaches to the use of digital data during a series of workshops that aimed to foster research collaborations to address contemporary complex social issues. The workshops involved researchers from across a large Australian university, public research institutions, and personnel working in policy and service settings. This is not uncommon in research addressing socio-technical issues where collaborators have varying experience in research involving human subjects (Zevenbergen et al. 2015). In our workshops, disagreement quickly surfaced in views of what constituted ethical risks and how the principles of research ethics should be applied to digital data in ways that presented potential barriers to developing and sustaining interdisciplinary research projects. We decided to facilitate a second series of structured workshop discussions to tease out and explore ethical issues associated with using digital data. These workshops would be used to consider emerging issues and whether they were adequately addressed in current national and international ethical guidelines. Participants would also collectively consider how these ethical issues could best be addressed.²

The second round of workshops involved university-based researchers and other research stakeholders. Drawing on these discussions, and relevant material including available guidelines and scholarly discourse, we identified five broad categories of ethical issues: consent; privacy and confidentiality; ownership and authorship; governance and custodianship; and data sharing and assessing the social benefits of research. These categories have some overlap with those identified by other scholars who are also grappling with the ethical challenges presented by digital data (Shilton and Sayles 2016). Our aim in presenting our findings is to contribute to ongoing debates of these issues and to develop some practical albeit provisional, guidance in navigating issues that accounts for diverse research contexts. Our insights are necessarily provisional because, as others

have noted, the potential uses of digital data for research are evolving in ways that risk leaving researchers with a lack of ‘ethical parameters for scientific research’ in this space (Metcalf and Crawford 2016, p. 2). Following an explanation of the process for the workshops, we discuss the five categories of issues that emerged from these discussions that elicited a practice-based approach for identifying and responding to emerging ethical issues. We conclude by offering key recommendations for developing guiding principles and suggest next steps for researchers and ethics committees.

Methods: running the workshops

In 2013 a cross-disciplinary and cross-sectoral forum engaging community and government partners was held, to foster interdisciplinary approaches to complex social challenges. At an initial workshop, two key issues became apparent. First, it was evident that there was lack of consensus among researchers on ethical issues related to the use of digital data for research. Second, it was recognised that technical capacities to generate, replicate and repurpose personal information as forms of digital data were advancing beyond accepted understanding of ethical issues associated with research.

Two subsequent workshops were held to focus specifically on ethical issues and with aims of developing some consensus on key ethical issues, challenges and strategies for addressing emerging risks. These workshops were convened in Melbourne in 2014, and involved researchers from a broad range of fields including computing and information systems, geomatics, engineering, mathematics, law, population health, e-health and e-learning, sociology and epidemiologic modelling. The workshops were designed and convened by an interdisciplinary team of academics with expertise in research ethics and experience on university ethics committees. This core group came from diverse fields including, biomedical research, privacy law, social science research, digital learning and human computer interaction and while some had been involved in the previous workshops, others were recruited to bring specific expertise into the team. Fifteen other workshop participants were senior academics and researchers working in universities and public research institutions in Melbourne, including representatives from the Victorian Health Promotion Foundation (VicHealth), the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and the City of Melbourne. These participants had long-term interests in digital technology, were actively involved in academic research and two had experience serving on ethics committees.

The workshops aimed to foster iterative and dialogic discussion among participants with diverse disciplinary expertise and understanding of research ethics. Encouraging reflexive consideration and exchange, this approach follows

¹ There are also the practical difficulties of adopting different processes and practices for managing sensitive digital data through adhering to specific guidelines such as ISO27001/2, a set of information security guidelines used to certify secure data practices. Digital data governance systems require careful development and implementation to ensure data security, and maintain participant confidentiality.

² The initial workshop “Building an evidence base to support planning for socially connected, equitable communities” was supported by the Carlton Connect Initiatives Fund. Five of the authors of this current article were involved in the organisation of the first workshop and the others were participants in subsequent workshops.

other efforts to foster interdisciplinary conversations exploring reasoning strategies that are used to consider ethical issues and dilemmas (Zevenbergen 2016; Zevenbergen et al. 2015). We also adapted a method that had been developed to explore ethical issues associated with the use of visual data (Waycott et al. 2015). Participants were asked to prepare for the workshops by writing short statements identifying critical and outstanding ethical issues they had encountered or observed in their research work. The issues that were identified were summarised and explored in workshop sessions and these discussions were recorded on prepared templates and further reflections were received via email. Supplementary interviews (N = 14) were also conducted with key informants with specific and relevant experiences or expertise in using or generating digital data, as members of ethics committees evaluating proposals involving the use of digital data, or as data custodians with an interest in the confidentiality and management of data. These informants were identified using a 'snowball sampling' strategy with workshop participants. The discussions from the workshops and the interviews were recorded or captured as text responses and coded to identify key sets of issues. A second workshop discussion focussed on insights obtained from the first workshop and identified research contexts and additional key issues. Insights and other material informed the development of guidelines to provide researchers and ethics committee members with practical advice on ethical issues relevant to the use of digital data in research. Drafts of the guidelines were extensively circulated among workshop participants and other experts for comment and feedback and were finalised and published in 2015 (Clark et al. (2015)).

Analysis of discussions in the workshops and interviews identified five key categories of ethical issues and the critical qualities of digital data that raise specific ethical issues. These insights have local and wider relevance and are presented with aims of contributing to other work and available resources.

Findings: key sets of ethical issues

Context

The guidelines addressed a significant gap. Currently, the Australian guidelines for ethical conduct in research (NHMRC 2007 [2014]) do not provide specific guidance for managing issues associated with internet-based research. A number of national and professional organisations such as the Association of Internet Researchers (AoIR), the American Psychological Association (Kraut et al. 2004), the Norwegian National Research Ethics Committee, (The National Committee for Research Ethics in the Social Sciences and the Humanities 2014), the European Union

(ALLEA Permanent Working Group on Science and Ethics 2011 (Revised 2017)) and computer scientists researching human computer interactions (McMillan et al. 2013) have published useful ethical guidelines, advice and suggestions for researchers using the internet (Ess and AoIR Ethics Working Committee 2002; Markham and Buchanan 2012; Utrecht Data School 2018). In some countries national research bodies have published guidelines specifically for researchers using the internet as a source of data and there are international data sharing agreements in some areas (Harriman and Patel 2014; Knoppers 2014; Knoppers et al. 2011). Nonetheless, national and discipline-inclusive ethical guidelines addressing the varied issues arising through the increasing use of digital data in human research remain rare.

Key ethical issues for research involving digital data

Participants noted a range of priority issues, including increasing capacities to harvest personal data for use in research and the ease with which these data can be replicated, circulated and used in ways that are likely not to be envisaged by the participants who provided their personal information, or researchers who collected forms of primary data that may be accessed for secondary analyses. The five categories of ethical issues that were distilled from the workshop discussions, interviews and available resources are presented below. Some of these issues are familiar to all research involving human participants, while others address the specific qualities of digital data. Our discussion of these issues draws attention to the ways in which ethical issues are being reformulated in contexts of digitalisation and expanding areas of research generating diverse forms of digital data. Researchers have responsibilities to understand emerging ethical issues and implement strategies to minimise risks for people whose experiences are the focus of research.

Consent

Although the conditions for consent are well established in research practice, there are concerns about gaining consent online that are specific to the contexts of internet-based research. Gaining informed consent in online environments is likely to require different processes from those used for gaining informed consent in face-to-face research encounters. In the latter, researchers have some capacity to assess whether participants understand the information they are being given and can provide additional explanation should it be required. In online environments, different processes can be used to gain informed consent. For instance, consent may be constituted through a registration process. Researchers, however, cannot be sure that participants have read and understood information they are given about the research and its ethical implications, including issues of privacy and

confidentiality, and how data will be used, stored or made available for other analyses (Convery and Cox 2012; Kanuka and Anderson 2008; Zimmer 2010). These challenges have led some researchers to declare that informed consent, normally considered the gold standard of ethical research, may in fact be “unnecessary” and could even mislead participants (Brown et al. 2016, p. 852). Brown et al. (2016) argue that consent (particularly in the form of “I agree” to the Terms of Service) may be no more than a legal protection for providers, “In interventionist HCI studies of this type, where the deployment of an application, service, or device and the access to it is the compensation for participation, the act of using the system could be interpreted as consent” (Brown et al. 2016, p. 856). In these authors’ opinion this form of consent should be separated out from researchers’ ethical responsibility to not harm the participants and it should not always be considered as the hallmark of ethical research. Ethics committees emphasise the importance of ensuring the availability of clear information about the research and ensuring that not only are there options for granting consent but also for participants to subsequently withdraw their information or opt out of research.

Researchers are also analysing material and content posted on social media, including social networking sites, and video-sharing sites, blogs and microblogs, as sources of research data (See Andalibi et al. 2016; Keim-Malpass et al. 2014; Liu et al. 2013; Vitak et al. 2015, 2016). For researchers analysing these forms of information it may be difficult or even impossible to establish consent. Markham and Buchanan make a distinction between gaining informed consent at the beginning of a research project and gaining consent if direct quotes and/or use of visual materials require individual consent the publication phase (Markham et al. 2012). Discussion of dynamic consent is a further development of consent for projects using digital data. Individuals may express preferences about how data is used in secondary research projects (Kaye et al. 2015).

Accepted processes for gaining consent to collect personal data for research are complicated by capacities for digital data to be repurposed for new and unanticipated analyses into the future. Individuals may be largely unaware that they are generating digital data that can be collected and analysed for research purposes (e.g. information emitted via mobile phones). In these circumstances, researchers may need to alert users that material is being collected for research purposes. Following the ‘Facebook study’ controversy, academic and public reaction to the case demonstrated a lack of consensus over issues of whether research involving social media should require gaining active forms of consent from users. Discussions also noted that expectations of gaining informed consent may be unrealistic in an age when people are routinely required to ‘consent’ to bewildering terms and conditions in order to access services. Alternative strategies

could involve using representative independent proxies to consider and assess ethical issues of particular projects or ensuring that users have timely opportunities to learn about any research that is being conducted or access to debriefing (Kahn et al. 2014; Voosen 2014).

Ethical issues related to consent are also relevant to the ways in which ‘big data’ and social media content are being analysed to predict future personal scenarios. Personal data are being analysed to identify situations that individuals themselves are yet to be aware of, for instance, research exploring the likelihood of new mothers experiencing post-natal depression based on their twitter posts (De Choudhury et al. 2013). These uses of digital data are fuelling ongoing debate as to whether it is ethical for researchers to collect material from social media as sources of data without informants’ knowledge or consent and, if not, what kinds of processes for gaining informed consent should be implemented.

Privacy and confidentiality

Protecting privacy in research using digital data is complicated by the range of data that might be collected. Many forms of personal information that are routinely collected, stored and reused are being digitalised and available to researchers and others. The benefits and risks of pervasive data collection from digital devices and the lack of public awareness of the ways in which digital data are being collected and used present new kinds of ethical issues in protecting privacy and confidentiality of research participants (Shilton 2009, 2010; Shilton and Estrin 2012). Increasingly, individuals are also putting self-generated data into public domains where it can be ‘harvested’ for research without their knowledge. Individuals are likely to have varying expectations about what might happen to their data, dependent on what kind of information it is. Many may be unaware of the amount of personal information that can be gleaned from data that they release into public domains. For example, a tweet of 140 characters (or less) actually contains around 9 kb of metadata that can include the precise location where the tweet was made, the device type, the user profile, information pertaining to followers and other details. Such rich sources of information can be used for a range of purposes that are rarely envisaged by tweeters. Even the use of health data for which there may be a clear public health benefit is controversial (Iversen et al. 2006; Vayena et al. 2015). As Iversen et al. (2006) point out, using personal information without explicit consent is legally permitted in the U.K. provided certain safeguards are in place; however data controllers may not interpret the law in this way and impose additional access restrictions.

There is no doubt that the rapid increase in mobile phone uptake and smart phone capacities, the widespread use of

Facebook and other social media tools and the pervasiveness of digital devices producing data, have challenged social, regulatory and academic concepts of privacy (Duckham and Kulik 2006; Harvey 2013). While there is no community consensus about what constitutes privacy, there is growing consensus within the scientific community on the benefits of sharing data (Safran et al. 2007; Walport and Brest 2011). Data linking and large scale epidemiological studies have shown the capacity to provide important new insights about population health, including disease surveillance, health services utilisation and have enabled evidence-based responses to health and social problems (Holman et al. 2008). Nissenbaum's (2009) work on privacy as a form of "contextual integrity" has been used as a basis for developing guidelines. However there are tensions between recognising the cultural contexts of concepts of privacy and at the same time as developing frameworks that facilitate international collaborations.

While the terms privacy and confidentiality are frequently used interchangeably, confidentiality is an ethical duty associated with professional practices (including research), and privacy is a right to freedom from intrusion that is protected and regulated by law. In Australia there are federal and state laws governing the collection and use of individuals' personal information, including health data. Such privacy-related laws set out specific requirements for institutions, researchers (and others) who are handling personal data. Techniques such as digital data matching and data-mining tools can increase the risks of privacy breaches occurring (Culnane et al. 2017; O'Keefe and Connolly 2010). Of particular concern are privacy issues arising from the collection of data from digital devices which link individuals with geographic locations. Other aspects of privacy in relation to the collection of data from digital devices require the individual and institutional protection of privacy, especially location privacy. A further aspect of privacy is the invasion of privacy (negative effects of invasion of privacy include spam, personal harm, intrusive interferences). In Australia, a popular misconception about digital data and privacy is that if data are collected from a publicly available source, such as a webpage, then the Privacy Act does not apply. This is not quite true. Data collected in the public domain are also governed by the operation of the Australian Privacy Principles under the *Privacy Act 1988* (Culnane et al. 2017; O'Keefe and Connolly 2011). However, the right to privacy is not an absolute right. In some circumstances it must be weighed against the rights of others and against matters that benefit society as a whole. The conduct of medical research is one such circumstance. Medical research is important for providing information to help the community make decisions that impact on the health of individuals and populations. However, research must be carried out in such a way as to minimise the intrusion on people's privacy.

Genomics is a field where data-sharing is considered particularly sensitive and yet there is potentially great benefit for the re-use of data. The potential for identification of individual human subjects associated with genomic information has led to developments which will ensure secure data sharing environments controlled by access committees are established as research is reviewed (Barsh et al. 2015; Kaye 2012). A variety of privacy protecting data structures are being developed by internet-based researchers as a way of anonymising data before researchers are able to access it (Koene et al. 2015). One of a number of approaches to resolving ethical issues associated with data sharing is the establishment of a Code of Conduct for genomic and health related research "to enable effective and responsible sharing of genomic and clinical data and to catalyze data sharing projects that drive and demonstrate the value of data sharing" (Knoppers 2014, p. 1e).

By contrast with privacy, confidentiality is an ethical duty that only authorised persons should have access to information. In research, maintaining participants' confidentiality involves strategies for keeping research data secure, and ensuring that access will be restricted to authorised users by appropriate data governance strategies. Issues related to confidentiality are also discussed in the section, 'Governance and custodianship'.

Ownership and authorship

The ease with which digital data can be replicated and reused raises important ethical issues in regards to the authorship and ownership of data. There is currently little consensus on who has responsibility for data at different stages of research and at what point personal information that is collected and aggregated in a database can be considered the property of particular individuals. This issue is particularly relevant in projects where data are managed by institutions (Borgman et al. 2006; Cragin et al. 2010). Research by Wallis and Borgman (2011) suggests that there is little agreement among academics about whether digital data can be 'owned' and, if so, what constitutes ownership. Are data owned by participants, research funding bodies, the principal researcher, members of research teams who collected the data, the guardians of data, or data storage services? In the case of digital data emitted by a device, is the person who is using the device considered its author or is the owner the person who has collected the data for research purposes? (McMillan et al. 2013; Wallis and Borgman 2011). This lack of clarity about data ownership has serious implications in attributing responsibility for key research processes such as data synthesis and interpretation, data governance and providing feedback to participants (Barker and Powell 1997).

Ethical issues about authorship emerge in association with responsibility for the long-term storage of data and

when data sets are subsequently re-used for new analyses and data linkage. There are a variety of ways in which researchers should be accountable for data they have collected including: ensuring the protection of sensitive data, the quality of the documentation accompanying data, that access to data is limited to authorised users, and agreements for re-using data (Khatri and Brown 2010; Weitzner et al. 2008). Issues of authorship and ownership are further complicated when research data moves between the public and the private domains, particularly when it acquires commercial value. The commodification of data, when self-generated data are sold on as business intelligence (Shilton 2009; Vayena and Tasioulas 2013), presents particular ethical challenges that are complicated by public–private partnerships and when these data are re-used for research. Allocating and enforcing long-term responsibility for the protection of sensitive material and the long-term maintenance of data can be challenging (Vitak et al. 2016).

Governance and custodianship

Issues of data governance are critical for ensuring that research that involves collecting or accessing digital data, is publically trusted. The safeguards established in data governance systems need to fully respect and protect individual privacy (Koene et al. 2015). Academic institutions, industry, and government have yet to fully develop integrated end-to-end systems for data governance and custodianship of digital data, including institutional information protection systems and ways of managing accountability (Fisher and Fortmann 2010; Rosenbaum 2010). Safeguards for the protection of data using technological means are being developed, offering increasingly sophisticated methods of protecting individual privacy (Malin et al. 2013; O’Keefe and Rubin 2015). As O’Keefe and Rubin (2015) point out, many of the risks of identity disclosure associated with the use of digital data can be managed by restricted access to data (e.g. data centres) and alteration of data (for example, de-identification). These approaches for protecting privacy rely on widespread acceptance and use of data governance systems and processes, including technological solutions that establish secure repositories for storing digital data (O’Keefe and Rubin 2015).

Currently research ethics committees reviewing applications require that a researcher is nominated as being responsible for security storage, management and access to data. However, guidelines do not clearly specify how these expectations should be achieved in relation to the management of digital data. For researchers, one way to reduce uncertainty in meeting these obligations is to develop some agreed standards for data governance and a comprehensive approach to management. There is a need to develop policy

consensus concerning the management of digital data by governments, research institutions and groups of data-users. Discussions of the variety of pathways and the changing locations of health data over time (e.g. data may be collected in one space, then housed in a restricted access data enclave and then re-purposed for use in another study), suggest that different methods of governance will be needed to manage data at different times in its life-cycle (Malin et al. 2013).

There are a variety of technological approaches to the protection of confidentiality and privacy such as obfuscation of geo-spatial data, de-identification, statistical disclosure control and the use of remote analysis servers that have been successfully implemented in Australia (Duckham and Kulik 2005, 2006; O’Keefe and Rubin 2015; Sinnott et al. 2014). These forms of good data governance and technological approaches are not, however, uniformly applied by institutions and ethics committees, nor are they specified in Australia’s National Statement on Ethical Conduct in Human Research, originally published in 2007 and most recently updated in 2015 (National Health and Medical Research Council 2015). The governance of data is also ruled by international treaties and national laws and is subject to increasingly complex international constraints (Fisher and Fortmann 2010; Lee and Gostin 2009; Safran et al. 2007). For researchers, the key ethical concerns are establishing good data governance practices in order to ensure data security and the ability to protect participants’ privacy and confidentiality. Good governance of data underpins the sharing of data, a system which relies on public trust, and trust between researchers and institutions (Knoppers et al. 2011; Rosenbaum 2010).

There are emerging views that national bodies will need to take leadership in promoting standards which underpin the re-use, ethical storage and sharing of digital data (Geissbuhler et al. 2013). Ideally, international consensus on these issues would support productive global collaborations and the European Union is currently involving its member states in developing this kind of framework (see <http://satoriproject.eu>). Notably, this is envisioned to enhance researcher confidence in making digital data publically available. A range of secure data enclaves has been established for access to, and linkage of, sensitive (individual) data (Burton et al. 2015). However these often have access and usage hurdles that are prohibitive for many researchers.

Data-sharing: assessing the social benefits of research

Data sharing offers important benefits because it can reduce the burden of research in populations and maximise the value and outputs of research; however, it also presents some ethical risks. Foremost concerns are promoting good practice in maintaining and ensuring public access to digital databases that can be used for research. The management,

organisation, access and preservation of digital data are all vital to research integrity and represent the great challenges of the information age. Despite the ease of copying and the flow and mobility of digitised information that is an intrinsic feature of digital data, complex consortium arrangements about research and data management requirements can hinder the sharing of data. There are also growing concerns that public access to data is threatened by the commercialisation of data and new markets for 'data products' leading to calls by civil society advocates of open government data to ensure continuing public access to these databases (See Bates 2012).

Data sharing facilities such as the UK Data Archive which act as repositories, curators and caretakers of digital from social science and humanities research projects (<http://www.data-archive.ac.uk/about/archive>) and the National Institutes of Health repositories list (at https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html) are front runners in setting up managements systems for ethical data-sharing. Data sharing is well established in a few areas such as genomics, although there have been some recent high-profile analyses of the risks to privacy that are presented (See Gymrek et al. 2013; Homer et al. 2008). In principle, many researchers are supportive of data sharing; however, in competitive research environments the actual extent of data sharing is limited in practice (Borgman 2012). Currently there are no national frameworks or guidance about the sharing of data or ethical issues associated with data-matching and there have been calls for greater guidance on these complex matters for some 20 years (Safran et al. 2007). Science and technology institutions and research bodies world-wide have identified the need for improved data sharing regulations (National Research Council 2009; Wellcome Trust 2011). Information protection and clear agreements regarding accountability need to be developed for responsible management of digital data produced by research, especially research involving 'big data'.

Discussion

There is widespread agreement that digital data and internet-based research technologies are transforming the research landscape. This offers new directions for research and benefits in the quality of knowledge that is generated. These possibilities, however, raise new and unfamiliar kinds of ethical issues. In this paper, we have discussed the findings from a collaborative, interdisciplinary project that sought to build shared understanding of emerging ethical issues when using digital data for research. Key categories of ethical issues were identified: consent; privacy and confidentiality; ownership and authorship; governance and custodianship; and data sharing. The workshops used a dialogic approach

to encourage participants to reflect on, and exchange views and experiences. The discussions generated a diversity of concerns that were organised into key categories of ethical issues to assist researchers and others to identify and respond to them in practical ways.

The anticipated exponential use of digital data for internet-based research requires enhanced international ethical guidelines that promote shared understanding of issues and cooperative approaches for minimising ethical risks. We recognise that there are challenges and opportunities in advancing these goals, including diverse philosophical, sociocultural and legal traditions that have influenced normative understanding of ethical issues within nations. Facilitating international, interdisciplinary and intersectoral discussions are clear priorities. Vitak et al. (2016) suggest that these discussions are key to developing the kind of trust needed to underpin interdisciplinary internet-based research, and research spanning academic and industry settings. At institutional and policy levels, the establishment of formal consensus-building bodies and strategies to develop a shared understanding of ethical issues and ways to address them would offer a positive approach for developing both national and international consensus and guidelines. There is a need for national regulatory bodies, such as the National Health and Medical Research Council (in Australia) to oversee the provision of advice on the ethical conduct of research using digital data. Initial steps have been taken in the provision of resources about managing data and the protection of privacy (see <https://www.nhmrc.gov.au/health-ethics/ethical-issues-and-further-resources>) but there is an absence of specific resources and guidance for internet-based research.

In Australia, the *National Statement on Ethical Conduct in Human Research* originally published in 2007 and most recently updated in 2015 (National Health and Medical Research Council 2015) guides the conduct of research involving human participants and, like similar protocols, was initially developed to address the circumstances and risks of research involving human clinical trials.³ Updating these important documents is urgently required in order to effectively respond to emerging social technologies (such as the use of smart phones for data collection and the collection of data from social media). Although recent revisions provide advice on the conditions under which informed consent for use of personal information can be waived and for the negotiation of consent through opt-out procedures, we echo other calls for additional ethical guidance be developed and incorporated into the National Statement to address issues arising through internet-based research (for example, see the

³ This document is currently under review and it is anticipated that an updated version due out in 2018 will include more specific guidelines for working with digital data.

recently released publication *Ethical Issues in Research Into Alcohol and Other Drugs: An Issues Paper Exploring the Need for a Guidance Framework*, (NHMRC 2012).

Internationally, scholars, institutions, professional associations and special interest groups in the UK and the USA have called for specific guidance for researchers using digital data (Ess and AoIR Ethics Working Committee 2002; Jones 2011; Markham and Buchanan 2012; Metcalf and Crawford 2016; Shilton and Sayles 2016; Warrell and Jacobsen 2014). The Economic and Social Research Council (ESRC) in Britain has a framework that briefly refers to internet-based research and alludes to the need for ethics review committees to seek specialist expertise for research based on digital data. In the United States, the Secretary's Advisory Committee on Human Research Protections (SACHRP) has developed a document that offers guidance about ethical and regulatory considerations for internet-based researchers (Secretary's Advisory Committee on Human Research Protections (SACHRP) 2013). However, privacy laws and accepted practices for gaining consent vary between different international jurisdictions, and this document has only limited application in contexts outside the United States. Researchers using internet-based technologies for educational studies are also encountering difficulties in interpreting how to apply National Statements to content that is generated for on-line educational platforms as part of course requirements (Chang and Gray 2013). Media attention garnered in recent instances of controversial research is giving stronger impetus to calls for clearer ethical guidelines. Ideally, such guidelines would reflect international consensus because of the increasing irrelevance of national boundaries in global information flows (Harriman and Patel 2014), and are relevant to public and private sectors.

Inadequate guidelines leave researchers and research ethics committees floundering in terms of assessing and responding to ethical issues associated with the use of digital data. Issues of privacy, data encryption, storage, access and data governance are ethical challenges that are frequently cited as being of concern for ethics committees in the United States (Buchanan et al. 2011; Buchanan and Hvizdak 2009; Metcalf and Crawford 2016). Members of ethics committees are rarely offered training on emerging ethical issues associated with digital data and internet-based research. Ethics committee members are expressing concerns about their lack of expertise in assessing risks, and the effectiveness of precautionary measures to minimise risks (Buchanan et al. 2011; Buchanan and Ess 2009). Other studies have found that a majority of human research ethics committees members and researchers believed that there are specific ethical issues raised by internet based research and there is a strong need for specific evidence-based guidelines and training for conducting internet-based research (Anderson and Sieber 2009; Fry 2013, 2014; Shilton and Sayles 2016). There are

also growing needs for ethics committees to secure advice from people with relevant technical expertise to assess issues which may arise in specific online environments. (See (O'Keefe and Rubin 2015) on reducing risks of disclosure when releasing data for secondary use). It may be timely to consider introducing additional methodological and technical expertise onto ethics committees.

The issues raised by ethics committee members are important to address. Methodological innovation presents ethics committees with two critical concerns and this applies to the expanding use of digital data for research (Daly 1996). The first concern is that projects that present risks are approved because committee members do not have adequate understanding of the ethical issues that may be associated with these methodological possibilities. The second concern is that potentially worthy projects are not approved (Bruckman 2014). The consequence of a lack of clear guidance is that both ethics committees and researchers in Australia describe inconsistency and uncertainty in responses from ethics committees about the use of the current National Statement when responding to ethics applications for research using digital data and for internet-based projects (O'Keefe and Connolly 2010). The flip side of these issues are evident in concerns among researchers seeking to conduct secondary analyses of data who report difficulty accessing digital data through data registries and inconsistent decision-making by human research ethics committees in granting access to available databases.

Other research suggests that human research ethics committees are taking longer to review and approve internet-based research (Fry et al. 2014). Ethics committees are increasingly required to consider and interpret complex privacy legislation, including scenarios where individuals may not be aware that personal data are being used for the purposes of research. This leads to tendencies to adopt risk-averse approaches and decline approval for researchers to access personal data, such as health information, despite privacy legislation in Australia that permits the use of de-identified information for research. In these situations, opportunities for research with strong potential for public benefit may be lost. At the same time, there are ongoing concerns among consumers in regards to the uses of personal data that are collected without their consent (O'Keefe and Connolly 2011, p. 5). These issues underline needs for increased guidance for both research ethics committees and researchers to identify and implement 'best practices' responses to ethical risks.

Best practices

This is an emerging field and we have argued here that there are currently few available guides outlining 'best practices'. However, we can point to existing practices and strategies

that enhance ethical practice. For instance, researchers have emphasised the importance of increasing the ‘digital literacy’ of researchers and ethics committees and improved ‘transparency’ in decision-making processes (Vidak et al. 2016, p. 950). Transparency requires explanations of ‘how’ and ‘why’ data are being collected, and why particular kinds of safeguards are being used to protect the privacy and confidentiality of participants. Such explanations serve to promote understanding and mutual trust between ethics committees, researchers and participants and operationalise practical strategies for ethical research using digital data. This involves iterative processes in which stakeholders learn from each other (Clark et al. 2015, p. 21).

The need for internationally relevant guidelines is pressing given the nature of internet-based research in which collaborations and data frequently cross national boundaries. It is helpful to begin with existing guidelines such as the US Secretary’s Advisory Committee on Human Research Protections, and then consider how to widen the scope to promote international relevance. Issues such as data sharing, data management and data governance are aspects of research using ‘big data’ that remain challenging for researchers and must be specifically addressed. Currently, there are limited models for designing data depositories that facilitate long-term data management, including processes for identifying individuals responsible for data governance and decision-making regarding the subsequent re-purposing of available digital data. International guidelines and standards are required to ensure the responsible management of digital data and suggested processes for identifying and resolving issues associated with accessing data for secondary analyses.

Accordingly we recommend that guidelines addressing the array of issues discussed in this paper should be developed using participatory and dialogic processes involving diverse stakeholders including relevant national research bodies, human research ethics committees, researchers (university, public sector and industry based) who use digital data in their research and community representatives. Regular updating of guidelines will also be required to keep abreast of technological developments. The guidelines developed by Clark et al. 2015, drawing on data discussed in this paper, is an example of a foundation for building shared understanding of ethical principles and effective practices for ethical research involving digital data.

Research agenda

Further research is needed to explore understanding and decision-making processes of research involving digital data from the perspectives of ethics committee members, participants, and relevant institutions. With respect to the conduct of human research ethics committees, it is important to examine how they respond to research projects using

digital data. What are the kinds of concerns they have about the ethical implications of obtaining, managing and storing digital data? Little research has been done about how ethics committee members weigh up the risks and benefits of digital research; it is important to examine how they formulate decisions about what constitutes benefits and how the risks to privacy and confidentiality are evaluated.

Research examining participants’ understanding of the risks and benefits of digital and Internet-based research is important. We need to understand from those most likely to be affected what they know about the ethical issues in this area, the basis of their knowledge, and how they would decide whether or not to participate in digital research. This would help inform the development of guidelines that are meaningful and relevant for those who are likely to face the most significant consequences.

We also have much to learn from research institutions, corporations, government bodies and other organisations involved in the collection, curating and secondary use of digital data. Examining how they understand ethical issues in this area and better understanding of their practices, especially around secondary analyses, would be useful; we want to be able to benefit from the data collected, while ensuring that the necessary protections are in place.

Educational implications

Our findings suggest that researchers and human research ethics committees are in need of further education about the ethical issues associated with digital data. As with all new and innovative research approaches, there is a need to be open and trusting about identifying potential ethical issues and finding productive means of addressing these. This also applies to any education that is undertaken.

For researchers, we have already pointed to the need for relevant, national and international guidelines that can alert researchers to potential ethical issues. The use of real-life case studies where ethical concerns have arisen, used in combination with ethical guidelines, can enable researchers to work through ethical scenarios with reference to ways that these can be addressed. Incorporating trigger questions in these case studies can alert researchers to ethical issues that they may not have previously encountered or even considered. Being reflexive with respect to both methodological and ethical rigour can ensure that researchers are sensitised to potential ethical concerns when using digital data.

For ethics committees, it would be helpful to develop education programs where experienced digital data researchers inform ethics committees about their research, highlighting both the benefits and the risks. In their decision-making, ethics committees are required to determine that the benefits of the research outweigh any potential risks. To do this they need to be informed about both the risks and the benefits;

this requires both researchers and ethics committee members being open to the possibilities, rather than digital research being dismissed outright without due consideration. As ethics committees become more familiar with digital research, they will be in a better position to advise researchers on how to mitigate possible risks and harms.

There is often very little feedback provided to ethics committees from researchers about what happens once research is approved. Other than perfunctory progress reports, ethics committees do not usually have the opportunity to hear from researchers about what has worked well, what kinds of unanticipated issues arose and how they were resolved. This feedback is important for gaining improved insights into ethical issues associated with emerging methods and contexts for research.

We also suggest that educational resources be developed to train researchers and ethics committee members in the conduct and assessment of research projects that involve digital data and internet-based research projects. These resources could include relevant guidelines, research publications that demonstrate the rigour and validity of internet-based research, discussions of relevant ethical issues that have occurred in practice and how they were addressed, as well as best practice in data security and data governance in relation to the management of sensitive research data.

Conclusion

It is clear that the use of digital data and internet-based research technologies is transforming the research landscape at a rapid rate. This kind of research has the potential to offer many benefits, but with this, comes a number of ethical challenges, some of which we are aware of, and others that remain unconsidered. We have a responsibility to ensure that the ethical issues that arise with digital research are identified and addressed where possible. In an attempt to do this, we have presented the findings from a collaborative project that explored emerging ethical issues associated digital research. Our findings suggested a dialogic process is critical for conducting and advancing discussions of complex ethical issues, and five key issues were identified. However, much still needs to be done. Wider public discussion is needed about internet-based research including topics such as the negotiation of online consent, and the use of data for which consent has not been obtained. Improved digital literacy and public debate about the social benefits of the secondary use of health data would help to clarify the public concerns about the use of routinely collected digital data. This can be used to inform further development of international guidelines that are congruent with contemporary societal values and expectations.

Acknowledgements The authors would like to gratefully acknowledge the Carlton Connect Initiative Fund, University of Melbourne.

References

- ALLEA Permanent Working Group on Science and Ethics. (2011) (Revised 2017). *The European Code of Conduct for Research Integrity*. Berlin: ALLEA—All European Academies. Retrieved from https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf.
- Andalibi, N., Haimson, O. L., De Choudhury, M., & Forte, A. (2016). *Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity*. Paper presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.
- Anderson, J. Q., & Raine, L. (2014). *The internet of things will thrive by 2025*. New York: The Pew Internet Project.
- Anderson, J. E., & Sieber, J. E. (2009). The need for evidence-based research ethics. *The American Journal of Bioethics*, 9(11), 60–62.
- Barker, A., & Powell, R. A. (1997). Authorship. Guidelines exist on ownership of data and authorship in multicentre collaborations. *BMJ: British Medical Journal*, 314(7086), 1046.
- Barsh, G. S., Cooper, G. M., Copenhaver, G. P., Gibson, G., McCarthy, M. I., Tang, H., & Williams, S. M. (2015). PLOS genetics data sharing policy: In pursuit of functional utility. *PLoS Genetics*, 11(12), e1005716.
- Bates, J. (2012). “This is what modern deregulation looks like”: Co-optation and contestation in the shaping of the UK’s open government data initiative. *The Journal of Community Informatics*, 8(2), 1–20.
- Beddows, E. (2008). The methodological issues associated with internet-based research. *International Journal of Emerging Technologies and Society*, 6(2), 124–139.
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059–1078.
- Borgman, C., Wallis, J. C., & Enyedy, N. (2006). Building digital libraries for scientific data: An exploratory study of data practices in habitat ecology. In S. Kapidakis, C. Mazurek, & M. Werla (Eds.) *Research and advanced technology for digital libraries* (pp. 170–183). Berlin: Springer.
- Boyd, D. (2016). Untangling research and practice: What Facebook’s “emotional contagion” study teaches us. *Research Ethics*, 12(1), 4–13.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679.
- Brown, B., Weilenmann, A., McMillan, D., & Lampinen, A. (2016). *Five Provocations for Ethical HCI Research*. Paper presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.
- Bruckman, A. (2014). *Research ethics and HCI ways of knowing in HCI* (pp. 449–468). Berlin: Springer.
- Buchanan, E. A. (2004). *Readings in virtual research ethics: Issues and controversies*. Hershey: IGI Global.
- Buchanan, E. A. (2011). Internet research ethics: Past, present, and future. In: M. Consalvo & Ch. Ess (Eds.) *The handbook of internet studies* (pp. 83–108). Chichester: Wiley
- Buchanan E.A., & Ess, C. M. (2009). Internet research ethics and the institutional review board: Current practices and issues. *ACM SIGCAS Computers and Society*, 39(3), 43–49.

- Buchanan, E., & Hvizdak, E. E. (2009). Online survey tools: Ethical and methodological concerns of human research ethics committees. *Journal of Empirical Research on Human Research Ethics*, 4(2), 37–48.
- Buchanan, E., Aycock, J., Dexter, S., Dittrich, D., & Hvizdak, E. (2011). Computer science security research and human subjects: Emerging considerations for research ethics boards. *Journal of Empirical Research on Human Research Ethics*, 6(2), 71–83.
- Burton, P. R., Murtagh, M. J., Boyd, A., Williams, J. B., Dove, E. S., Wallace, S. E., Tassé, A., et al. (2015). Data safe havens in health research and healthcare. *Bioinformatics*, 31, 3241–3248.
- Calear, A. L., & Christensen, H. (2010). Review of internet-based prevention and treatment programs for anxiety and depression in children and adolescents. *Medical Journal of Australia*, 192(11), S12.
- Chang, R. L., & Gray, K. (2013). Ethics of research into learning and teaching with Web 2.0: Reflections on eight case studies. *Journal of Computing in Higher Education*, 25(3), 147–165.
- Christensen, H., Griffiths, K. M., & Jorm, A. F. (2004). Delivering interventions for depression by using the internet: Randomised controlled trial. *BMJ*, 328(7434), 265.
- Clark, K., Duckham, M., Guillemin, M., Hunter, A., McVernon, J., O’Keefe, C., et al. (2015). Guidelines for the Ethical Use of Digital Data in Human Research. Retrieved from <http://www.carltonconnect.com.au>.
- Convery, I., & Cox, D. (2012). A review of research ethics in internet-based research. *Practitioner Research in Higher Education*, 6(1), 50–57.
- Conway, M. (2014). Ethical issues in using Twitter for public health surveillance and research: Developing a taxonomy of ethical concepts from the research literature. *Journal of medical Internet research*, 16(12), e290.
- Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1926), 4023–4038.
- Culnane, C., Rubinstein, B. I., & Teague, V. (2017). Health Data in an Open World. arXiv preprint arXiv:1712.05627.
- Daly, J. (1996). *Ethical intersections: Health research, methods and researcher responsibility*. Boulder: Westview Press.
- De Choudhury, M., Counts, S., & Horvitz, E. (2013). *Predicting postpartum changes in emotion and behavior via social media*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
- Duckham, M., & Kulik, L. (2005). A formal model of obfuscation and negotiation for location privacy. In H. W. Gellersen, R. Want, A. Schmidt (Eds.) *Pervasive computing* (pp. 152–170). Berlin: Springer.
- Duckham, M., & Kulik, L. (2006). Location privacy and location-aware computing. *Dynamic & Mobile GIS: Investigating Change in Space and Time*, 3, 35–51.
- Ess, C. & AoIR Ethics Working Committee. (2002). Ethical decision-making and Internet research: Recommendations from the AoIR Ethics Working Committee. Retrieved from <http://www.aoir.org/reports/ethics.pdf>.
- Fisher, J. B., & Fortmann, L. (2010). Governing the data commons: Policy, practice, and the advancement of science. *Information & management*, 47(4), 237–245.
- Fry. (2013). *Researcher and HREC attitudes about the ethics of e-research in Australia*. Paper presented at the Emerging issues, controversies and innovating for change., Freemantle.
- Fry. (2014). *Addressing the ethics of health eResearch with human participants*. Paper presented at the Towards Unified Global Research, Melbourne.
- Fry, C., Spriggs, M., Arnold, M., & Pearce, C. (2014). Unresolved Ethical Challenges for the Australian Personally Controlled Electronic Health Record (PCEHR) System: Key Informant Interview Findings. *AJOB Empirical Bioethics*, 5(4), 30–36.
- Geissbuhler, A., Safran, C., Buchan, I., Bellazzi, R., Labkoff, S., Eilenberg, K., et al. (2013). Trustworthy reuse of health data: A transnational perspective. *International journal of medical informatics*, 82(1), 1–9.
- Gleibs, I. (2014). Turning virtual public spaces into laboratories: Thoughts on conducting online field studies using social network sites. *Analyses of Social Issues and Public Policy*.
- Griffiths, K. M., & Christensen, H. (2007). Internet-based mental health programs: A powerful tool in the rural medical kit. *Australian Journal of Rural Health*, 15(2), 81–87.
- Gustafson, D. L., & Woodworth, C. F. (2014). Methodological and ethical issues in research using social media: A metamodel of Human Papillomavirus vaccine studies. *BMC Medical Research Methodology*, 14(1), 1.
- Gymrek, M., McGuire, A. L., Golan, D., Halperin, E., & Erlich, Y. (2013). Identifying personal genomes by surname inference. *Science*, 339(6117), 321–324.
- Hargittai, E. (2015). Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 63–76.
- Harriman, S., & Patel, J. (2014). The ethics and editorial challenges of internet-based research. *BMC Medicine*, 12(1), 124.
- Harvey, F. (2013). To volunteer or to contribute locational information? Towards truth in labeling for crowdsourced geographic information. In D. Sui, S. Elwood, M. Goodchild (Eds.) *Crowdsourcing geographic knowledge* (pp. 31–42): Berlin: Springer.
- Henderson, T., Hutton, L., & McNeilly, S. (2012). *Ethics and online social network research—developing best practices*. Paper presented at the Proceedings of Proceedings of the 26th BCS Conference on Human Computer Interaction.
- Hine, C. (2005). Internet research and the sociology of cyber-social-scientific knowledge. *The Information Society*, 21(4), 239–248.
- Holman, C. D. A. J., Bass, J. A., Rosman, D. L., Smith, M. B., Semmens, J. B., Glasson, E. J., et al. (2008). A decade of data linkage in Western Australia: Strategic design, applications and benefits of the WA data linkage system. *Australian Health Review*, 32(4), 766–777.
- Homer, N., Szelling, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., et al. (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genetics* 4(8), e1000167.
- Hoser, B., & Nitschke, T. (2010). Questions on ethics for research in the virtually connected world. *Social Networks*, 32(3), 180–186. <http://www.hhs.gov/ohrp/sac/rp/commsecbytopic/index.html>.
- Iversen, A., Liddell, K., Fear, N., Hotopf, M., & Wessely, S. (2006). Consent, confidentiality, and the data protection act. *BMJ*, 332(7534), 165–169.
- Jones, C. (2011). Ethical issues in online research. Retrieved from <https://www.bera.ac.uk/wp-content/uploads/2014/03/Ethical-issues-in-online-research.pdf>.
- Jouhki, J., Lauk, E., Penttinen, M., Sormanen, N., & Uskali, T. (2016). Facebook’s emotional contagion experiment as a challenge to research ethics. *Media and Communication*. 4, 75–85.
- Kahn, J. P., Vayena, E., & Mastroianni, A. C. (2014). Opinion: Learning as we go: Lessons from the publication of Facebook’s social-computing research. *Proceedings of the National Academy of Sciences USA*, 111(38), 13677–13679.
- Kanuka, H., & Anderson, T. (2008). Ethical issues in qualitative e-learning research. *International Journal of Qualitative Methods*, 6(2), 20–39.
- Kaye, J. (2012). The tension between data sharing and the protection of privacy in genomics research. *Annual Review of Genomics and Human Genetics*, 13, 415.

- Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: A patient interface for twenty-first century research networks. *European Journal of Human Genetics*, 23(2), 141.
- Keim-Malpass, J., Steeves, R., & Kennedy, C. (2014). Internet ethnography: A review of methodological considerations for studying online illness blogs. *International Journal of Nursing Studies*, 51(12), 1686–1692.
- Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148–152.
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Thousand Oaks: Sage.
- Knoppers, B. M. (2014). Framework for responsible sharing of genomic and health-related data. *The HUGO Journal*, 8(1), 1.
- Knoppers, B. M., Harris, J. R., Tassé, A. M., Budin-Ljøsne, I., Kaye, J., Deschênes, M., & Ma'n, H. (2011). Towards a data sharing Code of Conduct for international genomic research. *Genome Medicine*, 3(7), 46.
- Koene, A., Perez, E., Carter, C. J., Statache, R., Adolphs, S., O'Malley, C., et al. (2015). *Research ethics and public trust, preconditions for continued growth of internet mediated research*. Paper presented at the 1st International Conference on Information System Security and Privacy (ICISSP).
- Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences USA*, 111, 8788–8790.
- Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014b). Experimental evidence of massive-scale emotional contagion through social networks (Verma. IM. Editorial Expression of Concern and Correction). *Proceedings of the National Academy of Sciences USA*, 111, 10779.
- Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. *American Psychologist*, 59(2), 105.
- Lane, J., & Schur, C. (2010). Balancing access to health data and privacy: A review of the issues and approaches for the future. *Health Services Research*, 45(5p2), 1456–1467.
- Lee, L. M., & Gostin, L. O. (2009). Ethical collection, storage, and use of public health data: A proposal for a national privacy protection. *JAMA*, 302(1), 82–84.
- Liu, L., Huh, J., Neogi, T., Inkpen, K., & Pratt, W. (2013). Health vlogger-viewer interaction in chronic illness management, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 49–58). ACM Press.
- Lunnay, B., Borlagdan, J., McNaughton, D., & Ward, P. (2014). Ethical use of social media to facilitate qualitative research. *Qualitative Health Research*. <https://doi.org/10.1177/1049732314549031>.
- Malin, B. A., El Emam, K., & O'Keefe, C. M. (2013). Biomedical data privacy: Problems, perspectives, and recent advances. *Journal of the American Medical Informatics Association*, 20(1), 2–6.
- Markham, A., & Buchanan, E. (2012). Ethical Decision-Making and Internet Research Recommendations from the AoIR Ethics Working Committee (Version 2.0). Retrieved from <http://aoir.org/reports/ethics2.pdf>, Association of Internet Researchers website.
- Markham, A., Buchanan, E., & Committee, A. E. W. (2012). Ethical decision-making and Internet research: Version 2.0. *Association of Internet Researchers*.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.
- McKee, R. (2013). Ethical issues in using social media for health and health care research. *Health Policy*, 110(2), 298–301.
- McMillan, D., Morrison, A., & Chalmers, M. (2013). *Categorised ethical guidelines for large scale mobile HCI*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
- Metcalfe, J., & Crawford, K. (2016). Where are Human Subjects in Big Data Research? The Emerging Ethics Divide. The Emerging Ethics Divide (May 14, 2016). Big Data and Society, Spring.
- National Health and Medical Research Council. (2015). *Guidelines Under Sect. 95 of the Privacy Act (1988)*. Retrieved from <https://www.nhmrc.gov.au/guidelines-publications/e72>.
- National Research Council. (2009). *Ensuring the integrity, accessibility, and stewardship of research data in the digital age*. Washington, D.C.: National Research Council.
- NHMRC. (2012). *Ethical Issues in Research into Alcohol and Other Drugs: an issues Paper Exploring the Need for a Guidance Framework*. Retrieved from <https://www.nhmrc.gov.au/guidelines/publications/ds12>.
- NHMRC. (2007 [2014]). National Statement on Ethical Conduct in Human Research. <https://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research>.
- Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Palo Alto: Stanford University Press.
- O'Keefe, C. M., & Connolly, C. (2011). Regulation and perception concerning the use of health data for research in Australia. *Electronic Journal of Health Informatics*, 6(2), e16.
- O'Keefe, C. M., & Connolly, C. J. (2010). Privacy and the use of health data for research. *Medical Journal of Australia*, 193(9), 537–541.
- O'Keefe, C. M., & Rubin, D. B. (2015). Individual privacy versus public good: Protecting confidentiality in health research. *Statistics in Medicine*, 34(23), 3081–3103.
- Onsrud, H. J. (1995). Identifying unethical conduct in the use of GIS. *Cartography and Geographic Information Systems*, 22(1), 90–97.
- Rosenbaum, S. (2010). Data governance and stewardship: Designing data stewardship entities and advancing data access. *Health Services Research*, 45(5p2), 1442–1455.
- Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C., & Detmer, D. E. (2007). Toward a national framework for the secondary use of health data: An American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association*, 14(1), 1–9.
- Secretary's Advisory Committee on Human Research Protections (SACHRP). (2013). *Considerations and recommendations concerning internet research and human subjects research regulations (Recommendations for Research with the Internet)*. Washington D.C.: US Government. Retrieved from <http://www.hhs.gov/ohrp/sachrp/index.html>.
- Selinger, E., & Hartzog, W. (2015). Facebook's emotional contagion study and the ethical problem of co-opted identity in mediated environments where users lack control. *Research Ethics*, 12, 35–43.
- Shilton, K. (2009). Four billion little brothers? Privacy, mobile phones, and ubiquitous data collection. *Communications of the ACM*, 52(11), 48–53.
- Shilton, K. (2010). Participatory sensing: Building empowering surveillance. *Surveillance & Society*, 8(2), 131–150.
- Shilton, K., & Estrin, D. (2012). *Ethical issues in participatory sensing. Ethics CORE*. Urbana: National Center for Professional & Research Ethics
- Shilton, K., & Sayles, S. (2016). "We Aren't All Going to Be on the Same Page about Ethics": *Ethical Practices and Challenges in Research on Digital and Social Media*. Paper presented at the 2016 49th Hawaii International Conference on System Sciences (HICSS).
- Sinnott, R., Bayliss, C., Bromage, A., Galang, G., Grazioli, G., Greenwood, P., et al. (2014). The Australia urban research gateway.

- Concurrency and Computation: Practice and Experience, 27, 358–375
- Sinnott, R., Chhetri, P., Gong, Y., Macaulay, A., & Voorsluys, W. (2015). *Privacy-preserving Data Linkage through Blind Geospatial Data Aggregation*. Paper presented at the IEEE International Symposium on Big Data Security on Cloud (BigDataSecurity 2015), New York.
- The National Committee for Research Ethics in the Social Sciences and the Humanities. (2014). *Ethical Guidelines for Internet Research*. Retrieved from <https://www.etikkom.no/en/ethical-guidelines-for-research/ethical-guidelines-for-internet-research/>.
- Utrecht Data School. (2018). Data Ethics Decision Aid, from <https://dataschool.nl/deda/deda-for-research/?lang=en>.
- Vayena, E., Salathé, M., Madoff, L. C., Brownstein, J. S., & Bourne, P. E. (2015). Ethical challenges of big data in public health. *PLoS Computational Biology*, 11(2), e1003904.
- Vayena, E., & Tasioulas, J. (2013). Adapting standards: Ethical oversight of participant-led health research. *PLoS Medicine*, 10(3), e1001402.
- Vitak, J., Blasiola, S., Patil, S., & Litt, E. (2015). Balancing audience and privacy tensions on social network sites: Strategies of highly engaged users. *International Journal of Communication*, 9, 20.
- Vitak, J., Shilton, K., & Ashktorab, Z. (2016). *Beyond the Belmont Principles: Ethical challenges, practices, and beliefs in the online data research community*. Paper presented at the Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing.
- Voosen, P. (2014). Big data scientists face ethical challenges after Facebook study. *The Chronicle of Higher Education*.
- Wallis, J. C., & Borgman, C. L. (2011). Who is responsible for data? An exploratory study of data authorship, ownership, and responsibility. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1–10.
- Walport, M., & Brest, P. (2011). Sharing research data to improve public health. *The Lancet*, 377(9765), 537–539.
- Warrell, J. G., & Jacobsen, M. (2014). Internet research ethics and the policy gap for ethical practice in online research settings. *The Canadian Journal of Higher Education*, 44(1), 22.
- Waycott, J., Guillemin, M., Warr, D. J., Cox, S., Drew, S., & Howell, C. (2015). Re/formulating ethical issues for visual research methods. *Visual Methodologies*, 3(2), 4–15.
- Weitzner, D. J., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., & Sussman, G. J. (2008). Information accountability. *Communications of the ACM*, 51(6), 82–87.
- Wellcome Trust. (2011). Sharing research data to improve public health: Full joint statement by funders of health research Retrieved from Wellcome Trust website: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_peda/documents/web_document/wtvm049648.pdf.
- Whitehead, L. C. (2007). Methodological and ethical issues in Internet-mediated research in the field of health: An integrated review of the literature. *Social Science & Medicine*, 65(4), 782–791.
- Zevenbergen, B. (2016). Networked Systems Ethics—Guidelines.
- Zevenbergen, B., Mittelstadt, B., Véliz, C., Detweiler, C., Cath, C., Savulescu, J., & Whittaker, M. (2015). Philosophy meets Internet Engineering: Ethics in Networked Systems Research. (GTC workshop outcomes paper).
- Zimmer, M. (2010). “But the data is already public”: On the ethics of research in Facebook. *Ethics and information technology*, 12(4), 313–325.
- Zimmer, M. (2018). Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity. *Social Media + Society*, 4(2), 2056305118768300.