GUIDELINES FOR THE ETHICAL USE OF DIGITAL DATA IN HUMAN RESEARCH

(Authors in alphabetical order)

Karin Clark,
Matt Duckham,
Marilys Guillemin,
Assunta Hunter,
Jodie McVernon,
Christine O'Keefe,
Cathy Pitkin,
Steven Prawer,
Richard Sinnott,
Deborah Warr &
Jenny Waycott

THE UNIVERSITY OF
MELBOURNE

Carlton
Connect
Initiative

© Karin Clark, Matt Duckham, Marilys Guillemin,
Assunta Hunter, Jodie McVernon, Christine O'Keefe, Cathy Pitkin, Steven Prawer, Richard Sinnott,
Deborah Warr, Jenny Waycott (Authors in alphabetical order).

ISBN 9780 7340 5104 2

First printed: March 2015

This work is joint copyright. It may be reproduced in whole or in part for study or training purposes subject to acknowledgement of the source and is not for commercial use or sale. Reproduction for other purposes or by other organisations requires the written permission of the authors.

Further hard copies can be obtained from:

Associate Professor Jodie McVernon

Modelling and Simulation Unit, Centre for Epidemiology and Biostatistics,
Melbourne School of Population and Global Health,
3/207 Bouverie St,
The University of Melbourne,
Victoria 3010
E: j.mcvernon@unimelb.edu.au

Electronic copies of the Ethical Guidelines for the Use of Digital Data are available at the Carlton Connect website: www.carltonconnect.com.au

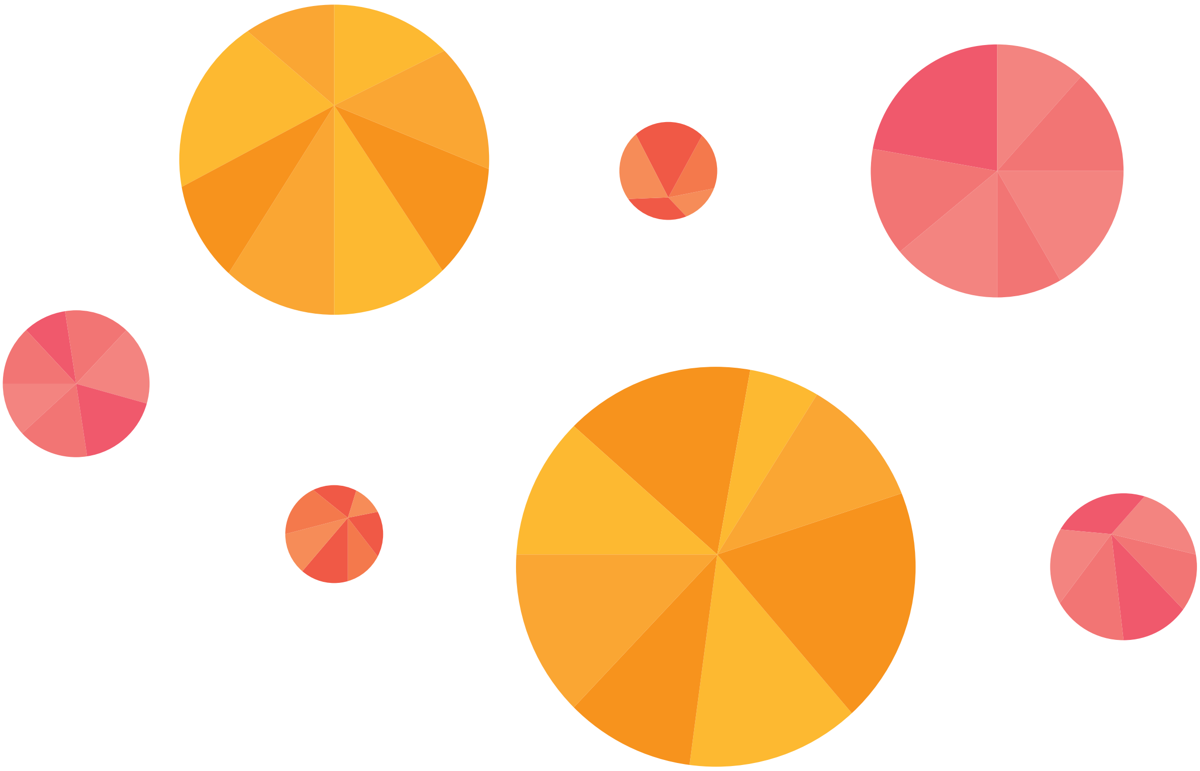
For citation:

Clark, K. Duckham, M. Guillemin, M. Hunter, A. McVernon, J. O'Keefe, C. Pitkin, C. Prawer, S. Sinnott, R. Warr, D. Waycott, J. (2015) *Guidelines for the Ethical use of Digital Data in Human Research*, The University of Melbourne, Melbourne.

GUIDELINES FOR THE ETHICAL USE OF DIGITAL DATA IN HUMAN RESEARCH

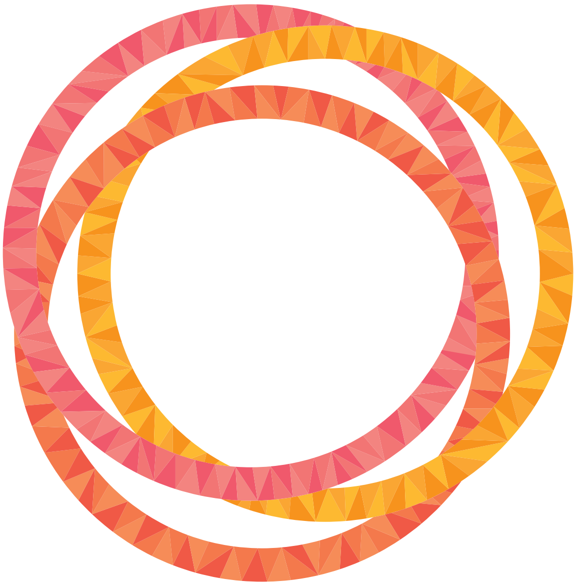
(In alphabetical order)

Karin Clark,
Matt Duckham,
Marilys Guillemin,
Assunta Hunter,
Jodie McVernon,
Christine O'Keefe,
Cathy Pitkin,
Steven Prawer,
Richard Sinnott,
Deborah Warr &
Jenny Waycott



CONTENTS

Authors: Karin Clark, Matt Duckham, Marilys Guillemin, Assunta Hunter, Jodie McVernon, Christine O'Keefe, Cathy Pitkin, Steven Prawer, Richard Sinnott, Deborah Warr, Jenny Waycott. (in alphabetical order)



Acknowledgements: We wish to thank participants in two Melbourne-based workshops, and academics engaged in individual interviews, for their contributions to these guidelines as part of the project Exploring the Ethical Frontiers Associated with the Use of Emerging Technologies in Social Research.

We further wish to acknowledge the contribution of Dr Susan Cox to Part C *Practical Approaches for Research Ethics Committees*, which is based on the corresponding section of the related *Guidelines for Ethical Visual Research Methods.*

We wish to acknowledge the initiative and diligence of our project officer, Dr Assunta Hunter, in assembling and synthesising this and other relevant information leading to the drafting of this document.

We are grateful for the useful and insightful comments on successive drafts made by colleagues at the University of Melbourne, City of Melbourne and CSIRO.

We would especially like to thank and acknowledge those people who were interviewed for their views and expertise as part of the project: (in alphabetical order)

Dr Rufus Black, Associate Professor Matt Duckham, Dr Serryn Eagleson, Mr Phillip Delaney, Associate Professor Craig Fry, Professor Marilys Guillemin, Professor Kim Lowell, Ms Delyth Lloyd, Dr Christine O'Keefe, Associate Professor Chris Pettit, Ms Cathy Pitkin, Professor Richard Sinnott, Professor Steven Prawer, Associate Professor Deborah Warr, Dr Jenny Waycott.

This project was supported by the Carlton Connect Initiatives Fund, University of Melbourne.

GUIDELINES FOR THE ETHICAL USE OF DIGITAL DATA IN HUMAN RESEARCH

The guidelines presented here have been developed to assist researchers who are conducting, and ethics committee members who are assessing, research involving digital data.

Digital data presents researchers and ethics committees with familiar and novel ethical issues. Accepted strategies for managing issues such as privacy and confidentiality, and informed consent, need rethinking. The qualities of digital data, including its mobility and replicability, present new kinds of ethical issues which emerge in relation to data governance, data security and data management.

This document has five parts.

Part A discusses key features of digital data and explains how these guidelines were developed.

Guidelines for researchers and human research ethics committees are presented in Parts B and C.

Part B addresses researchers and discusses five categories of key ethical issues and poses related guiding questions to consider when conducting research involving digital data:

- Consent
- Privacy and confidentiality
- Ownership and authorship
- Data governance and custodianship
- Data sharing: assessing the social benefits of research

Part C addresses members of human research ethics committees and provides guiding questions for reviewing projects involving the use of digital data.

The guidelines in Part B and C are formulated as discussions of key issues and arising questions. They are not intended to be prescriptive, but rather to contextualise and focus on key ethical risks in research involving digital data.

Part D is a glossary of key terms used in the document.

Part E lists resources that have informed the development of these guidelines and others which readers may find useful.

PART A: INTRODUCTION

These guidelines have been compiled to assist researchers and human research ethics committees to recognise and respond to ethical issues that are associated with the use of digital data in human research. While digital data have been around since the invention of computers, the growing range and availability of digital devices, the widespread use of social media, and the emergence of 'big data' are pushing researchers to reconsider their ethical responsibilities and obligations.

Digital data is the scientific term widely used to describe the kind of information that has been created in a computer mediated environment and which is transmitted as discrete information signals by the internet. Digital data are electronically produced and transmitted by the internet and may be captured or stored in data repositories.

Digital data are created in a variety of ways that concern researchers, and include:

- survey data from purposive collection of research data in online environments
- data routinely generated through contact with health professionals, hospitalisations, vaccinations and social service providers and increasingly through the use of direct-to-consumer services such as pathology and genetic testing services
- self-generated "lifelogging" data (including metadata) emitted from mobile phones and other "smart" appliances (e.g. Google Glass), generated through educational and lifestyle applications such as fitness monitoring devices and web-based games, gambling, dating, and posts on social media
- 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 about births, deaths, marriages; credit ratings, criminal convictions, immigration and customs records.

Digital data can be analysed and reanalysed in ways that may not be anticipated or expected by individuals whose personal information is part of the large data sets that are being produced. Facilitated by advances in digital technologies, our society has vastly increased capacities to create, store, replicate, exchange and extract data from data sets and bases. The re-purposing of data can include conducting secondary analyses on existing resources, and using techniques like data-mining and data linkage. Through these processes, linked data 'bundles' may be created and viewed as commodities to be bought and sold.

Prompted by the expanding scope of forms of digital data and technological capacities to manage data that are being collected and used for research, it is timely to revisit and reconsider ethical issues associated with research using digital data. These include new kinds of risks to privacy and to confidentiality, issues of consent and questions of when consent is required for researchers to access personal data. Ethical principles inform the requirement to show the social benefits of re-using and re-purposing data in subsequent analyses. These and other ethical issues are relevant for researchers and those involved in the construction and management of digital platforms, ethics committee members and research participants. These guidelines are orientated to these ethical issues and were produced through a consultative project involving researchers and others working in a range of disciplinary fields and organisations. The objectives were to consult with researchers across a broad range of fields to identify common ethical issues and strategies that are used to address ethical risks.

A key theme emerging from consultations is the need for institutions to establish their own policies and guidelines that specify institutional priorities regarding the use of digital data. These guidelines are offered as a broad framework for the kinds of questions which need to be addressed.

THE NEED FOR ETHICAL GUIDELINES FOR THE USE OF DIGITAL DATA

The rapid pace of technological change, and the increasingly interdisciplinary nature of research, present many challenges to researchers and ethics committee members grappling with, and understanding potential ethical risks in specific research projects. Further, researchers are concerned with navigating potential tensions between the need for innovation and commitment to a culture of responsible research. Researchers and ethics committee members are also required to identify ethical issues in fields where technology and information systems are constantly evolving (Fisher and Fortnum 2010, Brey, 2012).

BOX 1
EXCERPT FROM ASSOCIATION OF INTERNET RESEARCHERS ETHICS DOCUMENT (ESS ET AL 2002:3)

… both the great variety of human inter/actions observable online and the clear need to study these inter/actions in interdisciplinary ways have thus engaged researchers and scholars in disciplines beyond those traditionally involved in human subjects research: for example, researching the multiple uses of texts and graphics images in diverse Internet venues often benefits from approaches drawn from art history, literary studies, etc. This interdisciplinary approach to research leads, however, to a central ethical difficulty: the primary assumptions and guiding metaphors and analogies - and thus the resulting ethical codes - can vary sharply from discipline to discipline, especially as we shift from the social sciences (which tend to rely on medical models and law for human subjects' protections) to the humanities (which stress the agency and publicity of persons as artists and authors).

Complex ethical decisions arise in balancing the potential risks to research participants against the generalised benefits and cost effectiveness of research. Ethical issues are also relevant across all stages of the research process, including planning, data collection, dissemination of research findings, data storage and subsequent opportunities for data to be shared with other researchers, compiled into new databases and reanalysed to answer new research questions. Ethical decision-making increasingly needs to take into account current benefits and risks and consider future applications and analyses. Interdisciplinary research may require blending of culturally disparate views and in-depth interpretation of ethical issues.

In particular, digital data requires rethinking the adequacy of past practices to ensure ethical research. Harvesting digital data from social media and similar seemingly public online fora is fuelling debate defining notions of participation in research and about the applicability of traditional models for gaining informed consent. There are also concerns about the risks of identifiability in some research projects using digital data, and at the same time there are difficulties in authenticating online identities. Data linkage also presents risks to privacy and confidentiality. More broadly, there are emerging and pressing issues of ensuring data security, privacy and governance where there are multiple research partners who may be geographically dispersed, within and between countries.

Importantly, these guidelines aim to build on current efforts to address emerging ethical issues associated with the expansion of digital data. The

NHMRC *National Statement on Ethical Conduct in Research* (2007, updated March 2014) was drawn up at a time when research using digital data (such as online surveys, the use of smart phones for data collection and the collection of data from social media) was less common. Since then, capacities to collect and match large data sets are continuing to grow. The NHMRC acknowledges that its own guidelines are limited by the fact that it is impossible to foresee all the ways in which, into the future, data may be generated and what even constitutes the processes of human research. Notably, the NHMRC's recently released guidelines for Ethical Issues in Research into Alcohol and Other Drugs (2012) recommended that additional ethical guidance be developed and incorporated into the *NHMRC National Statement on Ethical Conduct in Research* to address emerging issues associated with the use of digital data for research.

HOW THESE GUIDELINES WERE DEVELOPED

In 2013 ethical and privacy concerns related to use of digital data emerged as a key theme among researchers participating in a forum held at the University of Melbourne that aimed to foster interdisciplinary responses to critical social challenges. Subsequently during 2014 two workshops were convened to specifically explore these issues with participants representing a range of academic fields including computer and information systems, law, mathematics, engineering, population health, ehealth and elearning, sociology and computer modelling in medicine. Academics and researchers working outside the academy were also involved in these workshops (VicHealth, CSIRO and the City of Melbourne were represented).

The first workshop was held in April 2014 and focused on the ethical challenges posed by using digital data generated through the use of devices. The second focused discussion held in September 2014 explored these key ethical issues in more detail. Working in small groups participants addressed case study examples highlighting key ethical issues. Discussions and notes from the workshops were written up and circulated to participants for comment and feedback. Workshop participants and a wider group of colleagues who were identified as having an interest in digital data, were also subsequently consulted one-on-one to elucidate key issues.

In addition, desk research was conducted to identify relevant articles, books and resources. These included human research ethics guidelines, and field guidelines from relevant professional associations that specifically addressed the use of digital data in internet-based research. Parts B and C of this report are informed by this material. Part B addresses researchers and discusses five key categories of ethical issues that were identified as critical for researchers to consider and understand when using digital data. Our discussions of these categories of ethical issues in the guidelines are supplemented by or illustrated using selected extracts from material we have collected. Guiding questions are offered to alert researchers to potential issues that need to be considered to enhance the ethical practice of research. Part C addresses members of human research ethics committees who assess projects using digital data. This is followed by a glossary (Part D) and a list of resources (Part E).

The guidelines are a work in progress rather than a definitive set of prescriptions. Ethical and rigorous research depends on how researchers conduct themselves in the actual practice of their research. In addition to researchers' own practices, ethical research is also positioned within specific institutional, disciplinary and national contexts. Finally, in such a rapidly evolving area, guidelines are inevitably provisional and require ongoing processes of reflexive practice and revision.

PART B: ETHICAL ISSUES FOR RESEARCHERS

Research ethics have been developed to ensure the equitable, just and respectful treatment of research participants and protect them from incurring harm through their involvement in research. Critical issues are avoiding physical and emotional harm; protecting research participants' privacy and confidentiality; and promoting research that serves the public good. Research ethics are grounded in four key principles:

- Respect for research participants is encapsulated in the accepted ethical statement of respecting the autonomy of participants and their right to be fully informed about the research endeavour;
- Research merit and integrity refers to the quality of the research processes in terms of meeting high standards of research practice, and the potential of the research to offer benefits. The conduct of the research must also be carried out in an ethical manner;
- Justice refers to the equitable and just treatment of research participants;
- Beneficence refers to minimising the risks of research and maximising the benefits of research.

In Australia's NHMRC National Statement on Ethical Conduct in Human Research (NHMRC 2007 updated 2014) research 'participants' refers to individuals who may or may not know that they are the subjects of research. These two positions are each relevant to research involving analyses of digital data as participants may or may not be aware that data pertaining to them are being used for the purposes of research. Researchers' ethical obligations also extend to individuals who are not participants in research but who may be affected by research activities when this impact is reasonably foreseeable. For example, when digital data are collected routinely, and generated in the course of transactions, individuals may be unaware that the data are being generated and may have no knowledge that the data may be used in research at a later date. However this may still need researchers' and ethics committees' consideration in terms of potential for harm.

Five key categories of ethical issues were identified as highly relevant to research using digital data. These issues are generally relevant to research; however, we argue that they require specific kinds of consideration when using digital data. Further, in identifying these five categories of issues we do not want to imply that these are the only ethical issues, but that these were deemed to be the most relevant. Nor do we want to suggest that these issues can be clearly separated out. Rather, the boundaries between research roles and the ethical issues which arise in working with digital data are sometimes blurred. The five categories of issues are:

- **CONSENT**
- **PRIVACY AND CONFIDENTIALITY**
- **OWNERSHIP AND AUTHORSHIP**
- **GOVERNANCE AND CUSTODIANSHIP**
- **DATA SHARING: ASSESSING THE SOCIAL BENEFITS OF RESEARCH**

CONSENT

Ensuring that participants are enabled to make informed decisions about their research participation is fundamental to consent in research. In consenting to participate in research, the process must be voluntary, and based on provision of sufficient information and adequate understanding of the purpose, aims and risks of the research, as well as what is required from participants. Although the conditions for consent are well established in research practice, there are issues regarding consent that are specific to using digital data.

While digital data continues to be collected through research projects it is important to ask whether participants are aware they are participating in research. Data may also be collected and analysed in ways that individuals who have contributed information or content remain largely unaware of. Accepted processes for gaining consent to collect personal data for research are being challenged by increasing capacities for digital data to be repurposed for new, previously unanticipated analyses. Increasingly, researchers are using material posted on social media, including social networking sites, audio-, photo- and video-sharing sites, blogs and microblogs, wikis, chat rooms, and virtual worlds as sources of research data. (e.g., Keim-Malpass et al, 2014; Liu et al, 2013).

Ethical aspects of 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 postnatal depression based on their twitter posts (de Choudhury et al, 2013). These uses are fuelling ongoing debate as to whether it is ethical for researchers to collect material from social media as sources of data and, if so, what kinds of processes for gaining informed consent should be implemented (See Box 3: Issues of consent and privacy in research using Facebook data).

Should consent be obtained, access to particular services may also be contingent on consent which poses a further ethical problem. The creation of online learning environments and the use of social media such as Facebook in teaching raises a series of specific ethical dilemmas for social researchers (Chang and Gray 2013). For example in online education environments there can be a requirement to complete certain tasks using social media. Social technologies are widely used in learning and teaching activities and may also be the basis for educational research (Henderson et al, 2013, Waycott et al 2010). Using data that individuals are obliged to provide in order to complete course assessment tasks, raises issues of consent. Research on students who are studying in such environments often aims to improve teaching and learning practices but can raise particular ethical challenges for the teacher/researcher:

- Should students be required to participate in social media (as part of student assessments), particularly if this is a form of interaction they are unfamiliar or uncomfortable with?
- If students are research participants, do they have the right to know the specifics of the research project and how information from the project (which may include work they have created and shared online) is to be used and stored?
- Are there conflicts of interest when the information students provide for an assessment task is also used as research data? In this case how can voluntary consent, without coercion, be ensured? Researchers may need to provide students with a choice to opt-out so they have the autonomy to choose whether or not their work is included in the research project.

According to the NHMRC National Statement on Ethical Conduct in Research, consent is currently required for all participants in research. The NHMRC National Statement grants exceptions to the use of medical data that have already been collected where gaining individual consent is unfeasible. In these circumstances, data can only be used if they are de-identified, that is, anonymised. Re-identifiability of data may have benefits for a person who receives a timely diagnosis or treatment as a result. Notably, recent cases have illustrated the reverse situation, that data thought to be irreversibly de-identified may be re-identified, even by a layperson with access to sufficient computer processing power (see Zimmmer 2010, Hayden 2013).

If deemed appropriate to gain informed consent, the process for doing this presents many challenges. Gaining informed consent in online environments involves different processes to those used for gaining informed consent in face-to-face encounters between researchers and potential participants. In face-to-face interactions, researchers have some capacity to assess whether participants understand the information they are being given and can provide additional explanation. In online environments, agreement to participate in research may be constituted through a registration process. However, researchers cannot be sure that participants have read and understood the information they have been given about the research. These include ethical implications, including issues of privacy and confidentiality and the ways in which data may be used, stored and made available for other analyses.

Possibilities for gaining informed consent to use content posted on social media need to be explored. Strategies need to consider the contexts in which information is being generated and shared (ie. the degree of public access in a chat room or in a social media platform). It may be possible to contact individuals through these platforms by sending them a Plain Language Statement and gaining consent to use material prior to it being posted or to collect information that has already been posted. It is widely agreed that obtaining consent from list-owners or moderators is insufficient, although they may be able to advise and assist researchers in implementing strategies that enable them to communicate with individuals using forums. For example, researchers could distribute information about their research and invite potential participants to contact them if they are interested in obtaining further information about the study.

Processes of gaining informed consent are further complicated in online environments where identity is not always clear. Individuals may have different or multiple online identities including some that are anonymised through the use of pseudonyms and avatars. Relationships between online identities, which offer individuals opportunities to construct alternate identities or to 'play' with identity, and real world contexts are not clear. For researchers, it may be difficult

BOX 2
'I AGREE' TO THE CONDITIONS OF USE

Participants may give consent to complex terms and conditions when providing information online and it is not clear whether they fully understand them. Is consent an 'implied contract' giving up personal data in exchange for a service ("a social contract") as seems to be the case in many digital environments? Is ticking 'I agree' when asked online enough? Many people are not necessarily aware of the full implications of pressing 'I agree' when asked on-line.

ODD SPOT: THE AGE 01.10.2014

In an experiment aimed at showing the disregard people have for privacy and online security requirements, some Londoners have agreed to give up their first-born child for free Wi-Fi. Customers were asked to agree to the condition as they logged-on to use free Wi-Fi at a central London cafe - and six people signed up.

to verify the identity of participants ('authentication of subjects'). Issues of consent, duty of care and harm in relation to research participants may be compounded by a lack of knowledge about who participants really are. Internet-based identification leaves open the fact that vulnerable, underage and in some cases un-defined populations may be unknowingly included in research studies (Kanuka and Anderson, 2008).

While research access to digital data is increasingly scrutinised, commercial repurposing of data is widespread and the use of social media by individuals provides public access to private information. Growing capacities to store, link and re-purpose data also raise critical questions about whether consent needs to be renegotiated in these circumstances. The re-use of data is increasingly common and there are concerns about data collected in one piece of research being used in a completely different piece of research without the explicit consent of the participants. Discussions about consent and re-negotiation of consent are linked to notions of authorship and ownership and to data governance and the ways in which researchers are accountable for their data.

BOX 3
ISSUES OF CONSENT AND PRIVACY IN RESEARCH USING FACEBOOK DATA

In 2008 a group of researchers released profile data that was collected from the Facebook accounts of an entire cohort of college students from a US university and published it as a data-set called 'Tastes, Ties and Time' (Zimmer, 2010; Lewis et al, 2008). Attempts were made to hide the identity of the institution and protect the privacy of the data subjects. However as soon as the research was published the source of the data was quickly identified. Researchers breached a number of ethical obligations including failing to gain specific consent from the students whose data were being harvested for the study. Researchers failed to ensure the students' expectations of privacy (even if information was posted on Facebook) and insufficient attention was paid to ensuring the efficacy of anonymisation techniques before the data were released (Zimmer, 2010). There were additional concerns that the institutional review board (research ethics committee) had allowed this research without explicit consent from the students and had not queried this aspect of the research protocol (perhaps due to their inexperience with internet-based research).

QUESTIONS FOR CONSIDERATION:

- Is an on-going process of informed consent (rather than a one-off consent) more appropriate for this research?
- Have all avenues for gaining informed consent from individuals to use potentially identifiable data been explored?
- Are participants aware that data collected for one research project may be reanalysed in future research projects?
- Is there a need for re-negotiating consent if the data are used by someone other than the researcher who collected it?
- Has consent been provided to link these data to other data (including personal data)?
- Does the consent process make clear the uses to which the population data (cf the individual data) may be put?
- When information is generated in one context, should consent be obtained to use this material for research purposes in another context?

PRIVACY AND CONFIDENTIALITY

Privacy and confidentiality are both key to ethical research practices. Privacy can be defined as the control that individuals have over who can access and manage their personal information. There are a number of kinds of privacy including location privacy and information privacy, both of which are substantially affected by the widespread use of digital devices and the production of digital data.

By contrast, confidentiality is the principle that only authorized persons should have access to information. In research, confidentiality refers to the process of keeping information gathered in research secure, and ensuring that access will be restricted to authorised users (data governance).

It is important to differentiate between the ethical value of confidentiality, which is a central aspect of the relationship between the researcher and research participants, and the legal definition of privacy. Confidentiality is one of the fundamental concerns of the *National Statement on Ethical Conduct in Human Research* (NHMRC 2007, updated 2014). In addition, privacy legislation focuses on data protection and controlling the uses that can be made of personal information, rather than on protecting the privacy of individuals in a broader sense. However, other laws (for example surveillance and listening devices laws such as the *Victorian Surveillance Devices Act 1999*) and the general law relating to breach of confidence may offer other remedies for the protection of privacy. These laws can be used to prevent breaches to confidentiality and the copying of digital data and its transmission to another person.

BOX 4
EXCERPT FROM THE NHMRC GUIDELINES OUTLINING THE RIGHT TO PRIVACY (2014:1)

An individual's right to privacy is a fundamental human right. This right is recognised in a number of international instruments, in particular, the *International Covenant on Civil and Political Rights (Article 17)* and the *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. Australia adopted the OECD Guidelines in 1984 and the principles in those guidelines were incorporated in the federal *Privacy Act 1988* (Privacy Act), which deals with personal information privacy protection, a component of the broader concept of privacy. However, the right to privacy is not an absolute right. In some circumstances, it must be weighed against the equally justified rights of others and against matters that benefit society as a whole.

The conduct of medical research presents one of these circumstances. Medical research is important for providing information to help the community make decisions that impact on the health of individuals and the community. However, it should be carried out in such a way as to minimise the intrusion on people's privacy. Optimally, this is done by obtaining the informed consent of participants prior to using their personal information. Where this is not practicable, de-identified information should be used. Where neither of these options is available, it may be that identified information needs to be used, even though consent of the individual or individuals has not been obtained, in order for the medical research to proceed.

In these latter cases, there is a need to balance the public interest in medical research against the public interest in privacy. These guidelines provide a framework in which such decisions can be made (*NHMRC Guidelines Under Section 95 of the Privacy Act 1988* revised March 2014).

Protecting privacy in research based on digital data is considerably more complex than protecting privacy in face-to-face research interactions. The benefits and risks of pervasive data collection from digital devices and the lack of public awareness of the use of digital data without explicit permission, are just two of the ethical issues posed by widespread use of digital data.

Privacy is one of those issues where the ethical challenges posed by digital data collected with consent are different to digital data which has been "harvested/used" without the individual's knowledge. Individuals have a wide variety of expectations about what might happen to their data and especially what might happen to their data without their consent. There is no community consensus about what constitutes privacy. The rapid increase in mobile phone uptake, the widespread use of Facebook and other social media and the pervasiveness of digital devices producing data, have shifted social, regulatory and academic concepts of privacy.

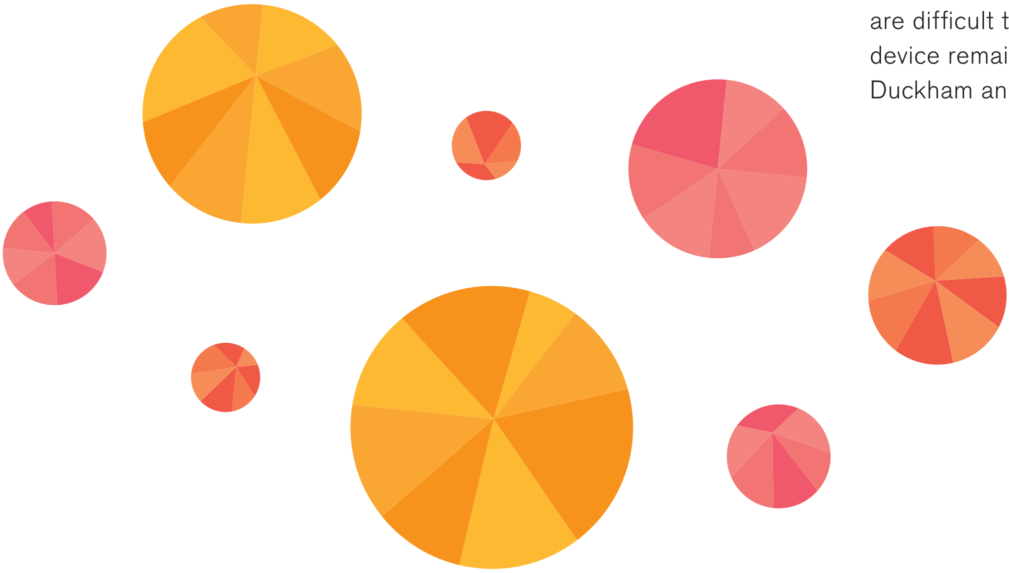
There are two main aspects of privacy in relation to the collection of data from digital devices:

1) Individual and institutional protection of privacy; and

2) Invasion of privacy (negative effects of invasion of privacy include spam, personal harm, intrusive interferences).

It is a popular misconception that data available in the public domain (eg on a webpage) is exempt from the Privacy Act. Data collected in the public domain is not exempt from the operation of the Australian Privacy Principles under the Privacy Act 1988 and the commercialisation of data has produced many examples of threats to privacy in the public domain (Shilton, 2009).

Privacy regulations world-wide refer to five fundamental principles of fair information practices (Duckham, 2015 forthcoming). Digital devices complicate some of these principles, for example the positioning data emitted by many digital devices is emitted as long as the device is turned on. Location information is used to keep the mobile phone user in contact with a network and is only secondarily available for other uses. Notice and transparency principles are difficult to maintain if the digital device remains turned on (See Duckham and Kulik, 2006).



BOX 5
THE FAIR INFORMATION PRACTICES

There exist around the world a wide range of regulations for dealing with private information. For example, more than 80 countries around the world have comprehensive laws that explicitly recognize information privacy rights (Greenleaf, 2012). While privacy regulations vary in their details, in general they all adhere to five fundamental principles of fair information practices (FIPs, also termed information privacy principles, IPPs). These five principles can be traced back to a 1974 U.S. Department of Health, Education, and Welfare report (on “Records, Computers, and the Rights of Citizens”):

1) Notice and transparency: Individuals must be made aware of when personal information about them is collected, by whom, and for what purpose.

2) Consent and use limitation: An individual’s consent is required in order to collect personal information about them. Personal information can only be collected for specified purposes, and the subsequent use of that information is limited to those purposes.

3) Access and participation: Individuals have the right to access personal data that refers to them. In case that stored data contains any inaccuracies, individuals may also require that errors be corrected.

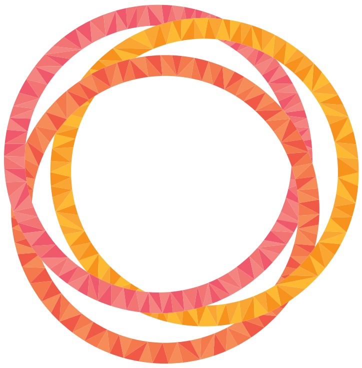
4) Integrity and security: Collectors of personal information must make reasonable efforts to ensure data is accurate and up-to-date. They must also protect against unauthorized access, disclosure, or use.

5) Enforcement and accountability: The collectors of personal information must be accountable for any failures to comply with the principles 1-4.

Duckham (2015 forthcoming) Confidentiality in the International Encyclopaedia of Geography Wiley AAP, New York.

QUESTIONS FOR CONSIDERATION:

- Does the data in question constitute personal information in the sense of the Privacy Act?
- Is there any mechanism, regulatory framework, or administrative structure that is designed to protect the individual’s privacy in relation to this project?
- Does the creation of data in this project challenge individual or community expectations about privacy?
- If explicit consent has not been obtained for this usage of data, does the public interest, as laid out in the NHMRC National Statement, support its use without consent?
- To what extent are the data gathered in this context considered personal and private, or public and available for research purposes?



AUTHORSHIP AND OWNERSHIP OF DIGITAL DATA

The whole arena of authorship and ownership of digital data is one where there is little consensus about who has responsibility for the data and at what point the individual has given up their right to control their personal data. This becomes an issue particularly in relation to data sharing and data management in projects where data is being re-used or shared (Wallis and Borgman, 2011).

Research among academics suggests that there is no agreement among researchers about whether digital data can be owned or indeed about who may be the owner (Wallis and Borgman 2011, Barker and Powell, 1997). Are data owned by the body that funded the research, the principal researcher, the research team, or the data storage service? In the case of digital data emitted by a device, is the person who is using the device considered its author? This lack of clarity about ownership of data has serious implications for consideration of responsibility for data synthesis and interpretation and feedback to participants.

BOX 6
AUTHORSHIP OF DATA: DISCLOSURE OF GENETIC INFORMATION

Whose responsibility is it to disclose genetic information if the discovery of this information is an unintended consequence of research?

There has been on-going controversy about the ethical responsibility of researchers to disclose genetic information that is discovered as a consequence of research, to participants in that research.

“Resolution of the question of whether there is a duty to return global or individual genetic research results depends on the type of study, the clinical significance and reliability of the information, and whether the study involves patients, genetically ‘at-risk’ families for a tested predisposition or healthy volunteers. Further confounding the emerging duty to return genetic research results is the situation in which the researcher is also a clinician and the participant is also a patient.” (Knoppers et al 2006).

Storing data in the ‘cloud’ or in a data repository does not necessarily change information ownership but it does change who is in control of access, who is charged with managing the data and perhaps who may be considered responsible for the data. Some issues of responsibility in relation to authorship can be resolved by clear guidelines written in the initial stage of collaborative research; these are part of good research practice and would include contributions to various stages of the research and to the analysis, design and conduct of the study. They can be used to outline responsibilities in terms of data management, storage, and destruction (where relevant), as well as contribution to research publication and dissemination plans.

The limitations of notions of authorship of digital data and potential ethical issues are apparent when it comes to re-using data sets, data linkage and the long-term storage of data. Wallis and Borgman (2011) suggest a variety of ways in which researchers can be accountable for their data including: protection of sensitive data, the quality of the documentation accompanying data, protection of data access, and support for data re-use.

The movement of research data between the public and the private domain may also be problematic in terms of authorship and ownership. Self-generated data which is then commercialised and used for business intelligence is a particular example (See Participant-Led Research Projects). Ethical issues can be complicated by public private partnerships and by data linkage (Zimmer, 2010). Unresolved questions arise about who has long-term responsibility for the quality of the data, the protection of sensitive material and the long-term maintenance of the data.

QUESTIONS FOR CONSIDERATION:

- What are the risks associated with the use of a data depository?
- Who has authority to access, release and manage this data?
- What processes have been used to anonymise this data?
- What potential harms may result from stripping data of identifiable information?
- Who is accountable for data quality, protection and access to data?
- Who is responsible for providing documentation and meta-data?
- Who is responsible for long-term maintenance of this data?
- Is data destruction (as a requirement of ethics applications) a relevant approach to digital data?

DATA GOVERNANCE AND CUSTODIANSHIP

The management, organisation, access and preservation of digital data are all vital to research integrity and represent great challenges of the information age. There is increasing emphasis on data access and preservation world-wide as digital data storage becomes more available and has become increasingly commercialised (Berman 2008). There are some issues which overlap authorship and governance domains. Data governance can be

BOX 7
PARTICIPANT-LED RESEARCH PROJECTS

A whole range of new applications enable users to track their health statistics through digital devices carried in their phones and in other wearable devices. They can monitor their blood pressure, heart rate, and other vital signs and they may use these devices to record their diet, medications and daily exercise patterns, including bike routes and running paths. The data emitted from these devices may also include their name, email details and information about work and leisure time activities. There are an increasing number of participant-led research projects using data generated from self-surveillance (Vayena and Tousoulis, 2013). Ethical concerns about the use of this data focus on "who collects the data, how it is handled, and what privacy protections are given" (Shilton, p. 49). Research that uses mobile phones to collect data may collect this data with the knowledge of the mobile phone owners, or the data may be "harvested" without the consent or knowledge of the person who generated it.

The information gathered through the mobile phone may be stored remotely, for example in the cloud, or with the phone manufacturers by the app that the phone owners have used. Concerns about these forms of storage include security and the extent to which the data becomes available to other parties. These digital data can be used and shared by commercial entities for research, financial gain and/or for advertising purposes. Companies such as health insurance companies are beginning to encourage their clients to use these self-tracking devices and are using the data produced to make insurance assessments.

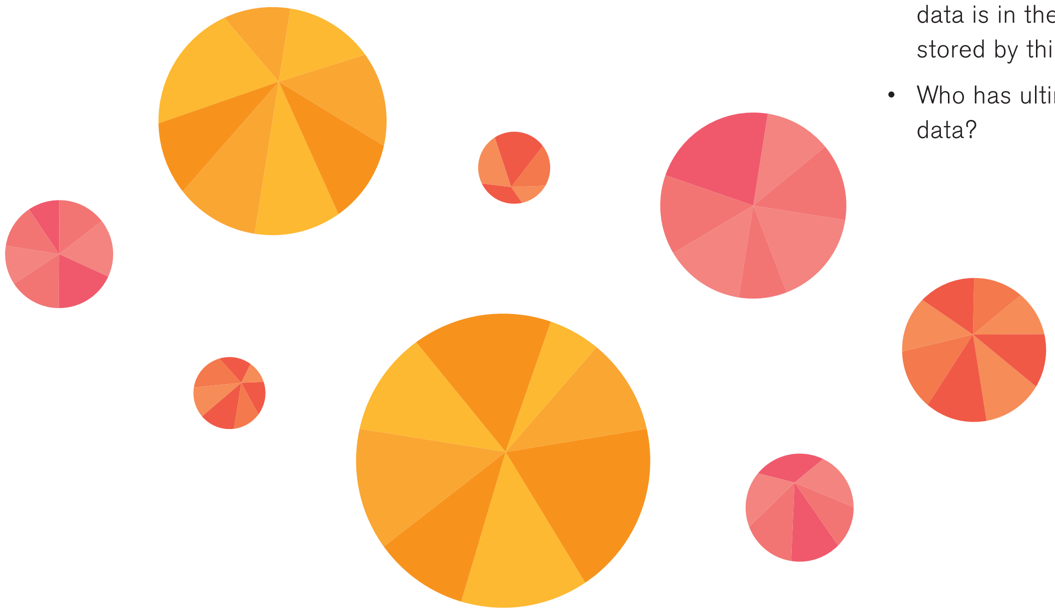
distinguished from authorship in that it deals with data storage and access to data and its possible re-use after the research has taken place.

Digitisation has enabled the generation of 'data products' that can be commercially sold. The ease of copying and the flow and mobility of digitised information is intrinsic to the nature of digital data. Complex consortium arrangements about research and the sharing of data can make these data management requirements difficult to put into practice. 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.

Research ethics applications need to ensure that there is someone responsible for storage, management and access to data. Policies need to be developed by institutions and groups of data-users about data management. Increasingly, the governance of data is also ruled by international treaties and national laws (Fisher and Fortnum 2010). At an institutional level a trade-off is often made between private storage networks (with associated costs) and publicly available tools and platforms that are expedient ways of managing data, but may lack adequate security checks and balances. For researchers the key ethical concerns are establishing good data governance practices in order to ensure data security and thus 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.

QUESTIONS FOR CONSIDERATION

- Are there processes in place to track the use of the data?
- Who is responsible for archiving data and/or deleting data if that is appropriate?
- Are processes in place to enable adequate data archiving and deletion as needed?
- How is access to data managed?
- What are the principles of data system management?
- How well informed and trained are the data gatekeepers?
- Is there a means of knowing when data has moved from one storage place to another or been copied/replicated in many places?
- Is there a way of retrieving data that has previously been shared?
- Who assumes responsibility once data is in the cloud or is managed/stored by third parties?
- Who has ultimate responsibility for data?



DATA SHARING: ASSESSING THE SOCIAL BENEFITS

Data re-use and data matching are techniques that have been enabled by the widespread creation and use of digital data and by increased computing capacities. The use of data from one research project by another research project is one form of re-purposing of digital data. Other forms of re-purposing of data include the use of administrative data-sets in research and the use of data coming from sources such as Twitter. Ethical challenges can arise when digital data produced by one project is used in another project or combined with data from another source, where such re-use must be approved or justified under the same framework as the original use of the data.

The cost-effectiveness of using existing data and making effective use of administrative and other publicly accessible databases provides a strong rationale for re-use of research data. Data can be re-purposed, representing efficient information use, but this process may carry additional risks to privacy and confidentiality beyond those present in the original use of the data. In the case of population health, the social benefits of timely and efficient access to health services data and public health research must be balanced against the risks of identity disclosure. There are recent attempts to establish codes of conduct for the sharing of data as a way of maximizing the social benefits of greater access to and use of data (Knoppers et al. 2011). Health research funding organisations have called for discussion of the values which underpin data-sharing and for the development of principles to guide on-going practice (Wellcome Trust 2011).

An ethical approach to data usage, the sharing of data and the re-purposing of data may require the implementation of an appropriate suite of protections from the variety of means whereby data can

BOX 8
DATA SHARING

International Data Sharing Code of Conduct

Preamble: This proposed International Data Sharing Code of Conduct seeks to promote greater access to, and use of, data in ways that are (as proposed by the joint statement by funders of health research):

- **Equitable:** any approach to the sharing of data should recognize and balance the needs of researchers who generate and use data, other analysts who might want to reuse those data, and communities and funders who expect health benefits to arise from research.
- **Ethical:** all data sharing should protect the privacy of individuals and the dignity of communities, while simultaneously respecting the imperative to improve public health through the most productive use of data.
- **Efficient:** any approach to data sharing should improve the quality and value of research and increase its contribution to improving public health. Approaches should be proportionate and build on existing practice and reduce unnecessary duplication and competition." (Knoppers et al 2011, p.46).

be made accessible for research while protecting the privacy and confidentiality of individuals (O'Keefe 2008). Technological approaches to data access which include the management of de-identified data by trusted agencies, the use of remote servers to enable restricted access to a dataset, and the use of confidentialised data are all developing (Lane and Schur 2010, O'Keefe 2008, p.8). The level of aggregation of data will influence the choice of protections. These means of accessing data are only some of a wide variety of available data management technologies, all of which have risks and weaknesses.

Data sharing and data repurposing highlight the need for researchers and ethics committees to understand the provenance of the data they are using. How was the data originally collected? Was any formal consent process involved? Do researchers also have a responsibility to evaluate whether further data usage should be aligned with the original purposes for which it was collected? In particular, is it likely that this secondary use may result in any harm to participants? What formal approval was sought for the use of the original data? Should the researcher assume that the same approval process applies to the re-purposed data?

Data collected for another purpose will necessarily have limitations. Are the data relevant and appropriate to the question under study? For example, while the re-use of data may be economically desirable, material taken from one social or economic context and used in another may not be directly applicable or referable to issues in the new setting. In considering likely social benefits of research, knowledge of the individual or agency seeking permission to re-purpose data, and their motivations for doing so, are clearly important. The re-purposing of digital data without participant consent has become increasingly common especially in the business domain. The use of re-purposed data in research is subject to the same approvals regime as the original data.

QUESTIONS FOR CONSIDERATION:

- Does the approval/permission regime for the original data include or preclude the new use of the data?
- Do researchers assessing data gathered in another context have a responsibility to understand the conditions of its original collection?
- Do researchers have a responsibility to assess whether the secondary use of the data is aligned with the original intent for which it was collected?
- Do researchers using data gathered by another research project have a responsibility to ensure that access to, and use of, the data does not pose a risk to individuals from whom it was originally collected?
- Is there a risk that in accessing the data collected by others that research participants will be adversely affected? How can this risk be evaluated?
- Do the benefits outweigh the potential risks and/or unintended consequences of repurposing data?
- What are the researchers' ethical and legal responsibilities in the use of re-purposed data?
- Is it possible to withdraw data from a project which may be secondary to the original research? (Is it ever possible to withdraw digital data?).

PART C: PRACTICAL APPROACHES FOR RESEARCH ETHICS COMMITTEES

It is important that ethics committees are aware and knowledgeable when reviewing applications involving emerging technologies. There are a variety of ways that ethics committees can engage with an ethics application involving digital data during the review and approval process. What follows are suggestions, organised in chronological order, to assist ethics committee members in their ethical review. Some of these steps may require further information from the researcher, or preferably, engagement in dialogue with the researcher about the project. The discussion of potential ethical issues and their resolution should be considered as a dialogic process.

The six areas addressed below are:

- acquiring necessary information about the proposed research
- asking appropriate questions of the researcher
- incorporating participant perspectives
- identifying strategies for handling ethical challenges
- learning from experience, and
- providing resources.

Some of the necessary information about the use of emerging technologies in the research project will be embedded in the research protocols.

The research ethics committee can draw on a list of preliminary questions for initial reviews of projects involving emerging technologies.

ACQUIRING NECESSARY INFORMATION ABOUT THE RESEARCH

1) Does this research involve:
    - Data collected through the use of digital devices?
    - Data collected in an internet mediated space?
    - Data that has been collected in one research project being linked or repurposed in another research project?
    - Data created/gathered by participants?

It will be important to consider the implications of ownership, authorship, and management of digital data both during the research process, and in the dissemination of findings and archiving of materials.

2) Will the data used or generated in the project:
    - Enable the possibility of re-considering or renegotiating consent?
    - Include information that could identify specific individuals or communities?
    - Include data which individuals might consider to be private or potentially harmful?

It may be necessary to describe specific risks or possible impacts of participation in the research that are related to the immediate use of data and on-going data collection.

ASKING APPROPRIATE QUESTIONS OF THE RESEARCHERS

It is important for ethics committees to understand how much knowledge and expertise the researcher has about the technology, devices or data being used or created in the research.

Some examples of appropriate questions to ask about the researcher include:

1) What level of expertise and/or experience does the researcher have with the creation and use of digital data?

2) How well does the researcher understand the conditions under which the data they are using has been created?

3) Is there a plan for the researcher to acquire necessary expertise in data management and data governance or to collaborate with others who have the necessary expertise?

4) Has the researcher demonstrated that they understand and can address the specific ethical issues that arise with the use of data collected through internet-based research, by digital devices or through the re-use of digital data produced in another context?

Technology has made the production of digital data in everyday life commonplace. However internet-based research methods and those using digital data may require increased levels of technical skills to manage and secure data. Researchers and participants also need to consider in what ways it is appropriate to seek permissions to participate in internet research and research using digital data, how to ensure that only those who have agreed to participate in the research are involved and, where relevant, how to authenticate the identities of those participating in research.

INCORPORATING PARTICIPANT PERSPECTIVES

How participants feel about what they are being asked to contribute is an important but often-overlooked aspect of research. It is perhaps especially important to inquire about this with internet based research, when data is collected online. This ensures that participants' preferences are respected and that they understand what is being asked of them.

Some examples of ways to incorporate participant perspectives include:

1) Ensuring that researchers ask participants whether they agree to the researcher returning to renegotiate consent should they consider using this data in another project.

2) Participants' level of understanding of the research project and agreement.

3) Providing an opportunity for participants to reflect on the experience of participating in the research:
    - Were there any unintended consequences of participating in the research (either positive or negative)?
    - Is there anything participants wanted to communicate to the researchers about participating in the study, both prior, during and following data collection?

Allowing for a reflective process and for participant feedback on the experience of research participation is important for researchers, for research ethics committees and for the larger research community. This is especially so with research methods that are new. It is one way to develop and share appropriate responses to ethical challenges. For researchers asking for feedback from participants is a way that they can gain insights into their research processes and how participants experience them.

The potential open-ness of digital data is shaping new models of research participation into so called "citizen science", which at a basic level may engage laypeople as owners and generators of their data. There are more advanced levels of participating in research which may involve participants shaping the research agenda collaboratively with trained scientists and analysing their own data. In this case, the participants whose data is being studied may actually be the ones to commission the research, organising this activity outside of a recognised research institution, and thus challenging many underlying assumptions of human ethics boards. Examples of this type of research participation are emerging in the life sciences. This calls for a slightly different perspective on engagement with research, and scrutiny by ethics committees.

IDENTIFYING STRATEGIES FOR HANDLING ETHICAL CHALLENGES

Rather than prevent research being undertaken, research ethics committee members can assist by providing suggestions on how to minimise the potential for harms.

For example some strategies for handling ethical challenges posed by the creation of digital data include:

1) Techniques used by researchers to protect privacy and confidentiality:
    - Restricted access to data (eg to approved individuals or projects)
    - Restriction of data through aggregation or perturbation
    - Provision of de-identified data
2) Techniques used for managing data governance:
    - Use of data enclaves
    - Use of data monitoring techniques such as specific access and user monitoring
    - Use of a data custodian and clear lines of responsibility for data
    - Identifying platforms that enable secure creation and storage of data

For many ethics committees there is a need for further education about the use of emerging technologies in research. Regular in-service education with a focus on the use of digital data and internet based research will help to educate ethics committees about these kinds of research. Further input about research using large data sets, the secondary use of data and data governance technologies would help to resource ethics committees effectively for the kinds of ethical challenges which may arise. Information about techniques used to minimise the risks to privacy and confidentiality and other kinds of ethical risks could be supplied to researchers and ethics committee members.

LEARNING FROM EXPERIENCE

Currently, there is very little feedback given to research ethics committee members about what happens with the studies they approve. It would be very helpful if outcomes were documented and all parties could learn from the challenges encountered with emerging research methods and especially those based on emerging technologies.

On completion of a study using novel methods:

- Researchers could write a study completion report documenting any ethically relevant challenges and outcomes
- Research ethics committee members could identify what had been learned in terms of how to assess applications using digital data in research.

These strategies would assist all stakeholders to learn from the research experience, to ensure that ethical and rigorous research can proceed.

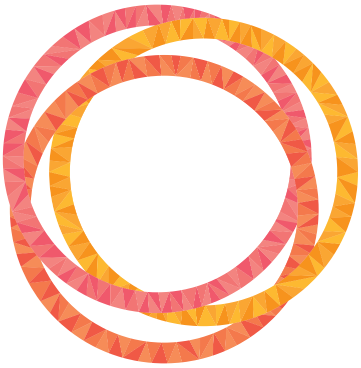
In the case of research using digital data it has been suggested that research ethics committees should be given some on-going training in ethical aspects of research using digital data.

PROVIDING RESOURCES

Research ethics committees need to have access to resources on the use of digital data in research and need to be able to provide these to others concerned with the ethical review of research.

Examples of necessary resources include:

- Key publications or guides on the use of digital data in research
- Articles about research using digital data that demonstrate an ethical approach to data collection and analysis
- Discussions of case studies which demonstrate the resolution of ethical issues
- Codes of ethics or practical guidelines being used by associations or researchers working with digital data and internet-based research (see resources section).



PART D GLOSSARY

Anonymity
The concealing of the identities of research participants in all documents resulting from the research.

Confidentiality
The ethical principle that only authorised persons should have access to information (encoded in regulations/codes of conduct). In research this refers to the process of keeping information gathered in research secure and ensuring that access will be restricted to authorised users and is part of data governance.

Consent
The process whereby participants in research agree to be part of the research process. Participants must be given sufficient information about the research project, its aims and outcomes, such that the researcher is confident that the participant can be described as having 'informed consent'.

Crowd sourcing
"[T]he act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the larger network of potential laborers" (Howe in Graber and Graber, 2013, p. 115).

Data collection within the public realm
This includes data collected without a person's direct consent (eg through mobile phone technology, electronic devices attached to clothes, geographical/spatial tracking data). Also data collected via registration (ie with consent) or with the individual's active participation (participant generated information).

Data enclaves
A data enclave is a secured environment in which data can be stored and accessed remotely (ie without direct access to the data). Access to information is controlled by data management personnel.

Data governance
"Data governance is defined as the process by which stewardship responsibilities are conceptualized and carried out, that is, the policies and approaches that enable stewardship. Data governance establishes the broad policies for access, management, and permissible uses of data; identifies the methods and procedures necessary to the stewardship process." (Rosenbaum 2010 p.1444-1445).

Data literacy
"With the advent of the personal computer and the web, information literacy requires both statistical literacy and data literacy. Students must be information literate: they must be able to think critically about concepts, claims and arguments: to read, interpret and evaluate information. Statistical literacy is an essential component of information literacy. Students must be statistically literate: they must be able to think critically about basic descriptive statistics. Analysing, interpreting and evaluating statistics as evidence is a special skill. And students must be data literate: they must be able to access, assess, manipulate, summarize, and present data. Data literacy is an essential component of both information literacy and statistical literacy." (Shield 2004 p. 4).

Datasets
Sets of data collected for a particular purpose.

Data re-purposing means using data previously created for one specific purpose which is then used for a completely different purpose.

Data re-use means using data from a project more than once for the same purpose.

De-anonymisation (or re-identification):
The linking of an individual's identity with a dataset record.

"Even if identifying information such as names, addresses, and Social Security numbers has been removed, the adversary can use contextual and back- ground knowledge, as well as cross-correlation with publicly available databases, to re-identify individual data records" (Narayan and Shmatikov 2013, p.1).

De-identified data
Data from which personal identifiers such as an individual's name, address, and date of birth have been removed. (Similar to anonymity).

Game research hybrids (gremes)
Crowd sourcing protocols that take the form of games. "Given that gremes rely on the draw of their game-like aspects, one can expect future crowdsourcing protocols to show significant improvement in the features of the game that attract and hold players' attention (Graber and Graber 2013, p.116).

Information privacy
This aspect of privacy is defined as the "claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" (Westin 1967, p.7).

Location privacy
This is a particular case of information privacy, concerned with the privacy of information about an individual's location.

Micro-data
Information about specific individuals.

Participant led research (PLR)
Communities of individuals engaged in establishing and conducting health research projects, often using self-generated data and frequently about specific medical conditions. The results of such participant-led research have already appeared in leading biomedical journals.

Participatory sensing
Participant gathering of data generated through mobile devices. This school of thought emphasises individual participation as one means by which personal data can be managed and privacy risks decreased.

Privacy Legislation (Australia) Commonwealth Privacy Act (1988)
"Privacy legislation in Australia includes: state and territory specific Privacy and Information Acts (various) in all states except except Western Australia. It also includes privacy and confidentiality provisions contained within other laws, research guidelines, codes of conduct and those contained in the common law."
(O'Keefe and Connolly 2011, p.3).

Re-purposing of data
Research data produced by one project may be used in another project or may be combined with data from another research project. Data can be re-used, re-discovered, or re-purposed. This may carry additional risks to privacy and confidentiality.

Responsible Research Innovation (RRI)
"The RRI programme creates an opportunity for reflection, where decisions about research goals are made not exclusively on the grounds of their technical or scientific attributes; so that, in addition to addressing technical grand challenges, RRI asks all stakeholders to consider the potential impacts, risks, and uncertainties of research outputs to wider society." (Stahl et al 2013, p.213).

Social networking sites
"Web-based services that enable individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate with a list of other users with whom they share a connection, and (3) view their list of connections and those made by others within the system" (Boyd and Ellison 2009, p.210).

PART E RESOURCES

BOOKS AND ARTICLES

Barker, A. & Powell, R. A. (1997). Authorship. Guidelines exist on ownership of data and authorship in multicentre collaborations. BMJ: British Medical Journal, 314, 1046.

Berman, F. (2008). Got data? A guide to data preservation in the information age. Communications of the ACM, 51(12), 50-56.

Boyd, D. M. & Ellison, N. B. (2008). Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication, 13, 210-230.

Buchanan, E. & Zimmer, M. (2012). Internet research ethics. In E.N. Zalta (ed). Stanford Encyclopaedia of Philosophy. Winter Edition.

Buchanan, E. A. (2011). Internet Research Ethics: Past, Present, and Future. The Handbook of Internet Studies. Chichester: Wiley-Blackwell.

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.

De Choudhury, M., Counts, S., & Horvitz, E. (2013). Predicting postpartum changes in emotion and behavior via social media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 3267-3276). Association for Computing Machinery Press, New York.

Duckham, M. & Kulik. L. (2006) Location privacy and location-aware computing. In Drummond, J., Billen, R., João, E. and Forrest, D. (eds) Dynamic & Mobile GIS: Investigating Change in Space and Time (pp. 35-51). CRC Press, Boca Raton, FL.

Duckham, M. (2015 forthcoming) Confidentiality. In Wiley-AAG International Encyclopaedia of Geography, New York.

Duncan, G., Elliot, M., & Salazar-Gonzalez, J.J. (2011). Statistical Confidentiality. Springer, New York.

Eysenbach, G. & Till, J. E. (2001). Ethical issues in qualitative research on internet communities. British Medical Journal, 323, 1103-1105.

Fisher, J. B. & Fortmann, L. (2010). Governing the data commons: Policy, practice, and the advancement of science. Information & Management, 47, 237-245.

Floridi, L. (ed) (2010). The Cambridge Handbook of Information and Computer Ethics. Cambridge University Press, Cambridge.

Fry, C. (2014). Addressing the ethics of health eResearch with human participants. Paper given at eResearch Australasia 2014: Towards Unified Global Research, 27-31 Oct, Melbourne.

Giannotti, F., & Pedreschi, D. (2008). Mobility, Data Mining and Privacy: Geographic Knowledge Discovery. Springer, Berlin.

Graber, M. A., & Graber, A. (2013). Internet-based crowdsourcing and research ethics: the case for IRB review. Journal of Medical Ethics, 39(2), 115-118.

Gray, K. (2008). Educational technology practitioner-research ethics. In M. Quigley. (ed) Encyclopedia of Information Ethics and Security (pp. 164–169). USA: Idea Group Inc, Hershey, PA.

Grunwald, A. (2011). Responsible innovation: bringing together technology assessment, applied ethics, and STS research. Enterprise and Work Innovation Studies, 7, 9–31.

Gubrium, A. C., Hill, A. L., & Flicker, S. (2013). A situated practice of ethics for participatory visual and digital methods in public health research and practice: a focus on digital storytelling. American Journal of Public Health, e1-e9.

Hayden, E. C. (2013). Privacy protections: The genome hacker. Yaniv Erlich shows how research participants can be identified from\'anonymous\'DNA. Nature, 497(7448), 172-174.

Henderson, M., Johnson, N. F., & Auld, G. (2013). Silences of ethical practice: dilemmas for researchers using social media. Educational Research and Evaluation 19(6), 546-560.

Hooley, T., Wellens, J., & Marriott, J. (2012). What is Online Research? Using the Internet for Social Science Research. A&C Black, London.

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E., Spicer, K., & de Wolf, P.P. (2012) Statistical Disclosure Control, John Wiley & Sons, United Kingdom.

Israel, M. (2014) Research Ethics and Integrity: Beyond Regulatory Compliance (2nd ed) Sage Publications, London.

Kanuka, H., & Anderson, T. (2008). Ethical issues in qualitative e-learning research. International Journal of Qualitative Methods, 6(2), 20-39.

Keim-Malpass, J., Steeves, R. H., & Kennedy, C. (2014). Internet ethnography: A review of methodological considerations for studying online illness blogs. International Journal of Nursing Studies, 51(12), 1686-1692.

Knoppers, B. M., Joly, Y., Simard, J., & Durocher, F. (2006). The emergence of an ethical duty to disclose genetic research results: international perspectives. European Journal of Human Genetics, 14(11), 1170-1178.

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, 46.

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, 1456-1467.

Lewis, K., Kaufman, J. & Christakis, N. (2008). The taste for privacy: An analysis of college student privacy settings in an online social network. Journal of Computer-mediated Communication, 14, 79-100.

McKee, R. (2013). Ethical issues in using social media for health and health care research. Health Policy, 110 (2), 298-301.

McKee, H. A., & Porter, J. E. (2009). The Ethics of Internet Research: A Rhetorical, Case-Based Process Peter Lang, New York.

Moreno, M., Fost, N., & Christakis, D. (2008). Research ethics in the Myspace era. Pediatrics, 121(1), 157–161.

National Research Council Panel (2007). Putting People on the Map: Protecting Confidentiality with Linked Social-spatial Data. National Academies Press, Washington.

Narayanan, A., & Shmatikov, V. (2008). How to break anonymity of the Netflix prize data set (pp. 111-125). In IEE Symposium on Security and Privacy, Oakland CA.

O'Keefe, C.M. (2008) Privacy and the use of health data - reducing disclosure risk. Electronic Journal of Health Informatics 3(1) e5.

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.

Onsrud, H. J. (1995). Identifying unethical conduct in the use of GIS. Cartography and Geographic Information Systems, 22, 90-97.

Owen, R., Bessant, J. & Heintz, M. (2013). Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society, John Wiley & Sons, New York.

Porter, J.E., & McKee, H.A. (2009) The Ethics of Internet Research: A Rhetorical, Case-Based Process. Oxford: Peter Lang.

Schield, M. (2004). Information literacy, statistical literacy and data literacy. IASSIST Quarterly, 28, 6-11.

Shilton, K. (2009). Four billion little brothers? Privacy, mobile phones, and ubiquitous data collection. Communications of the ACM, 52, 48-53.

Stahl, B. C. (2011). What does the future hold? A critical view of emerging information and communication technologies and their social consequences. In Chiasson, M., Henfridsson, O., Karsten, H., & De Gross J.I. (eds) Researching the Future in Information Systems: IFIP WG 8.2 Working Conference, Future IS 2011, Turku, Finland, June 6-8, 2011, Proceedings 1st ed (pp. 59-76). Heidelberg: Springer.

Stopczynski, A., Pietri, R., Pentland, A., Lazer, D., & Lehmann, S. (2014). Privacy in Sensor-Driven Human Data Collection: A Guide for Practitioners. http://arxiv.org/abs/1403.5299

Vayena, E., & Tasioulas, J. (2013). Adapting standards: Ethical oversight of participant-led health research. PLoS Medicine, 10(3), e1001402.

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.

Waycott, J., Gray, K., Clerehan, R., Hamilton, M., Richardson, J., Sheard, J., & Thompson, C. (2010). Implications for academic integrity of using web 2.0 for teaching, learning and assessment in higher education. International Journal for Educational Integrity, 6(2), 8–18.

Wellcome Trust. (2011). Sharing research data to improve public health: full joint statement by funders of health research. Available: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_peda/documents/web_document/wtvm049648.pdf.

Young, S. D., Holloway, I. W., & Swendeman, D. (2014). Incorporating guidelines for use of mobile technologies in health research and practice. International Health. Doi: 10.1093/inthealth/ihu019

Zimmer, M. (2010). But the data is already public: on the ethics of research in Facebook. Ethics and Information Technology, 12, 313-325.

PROJECTS

The Ethics of Biomedical Big Data: Oxford Internet Institute
http://www.oii.ox.ac.uk/research/projects/?id=130
A current (2014-2015) research project evaluating a European framework for the ethical use of big data in biomedical research.

The Framework for Responsible Research and Innovation in Information and Communication Technology (FRRIICT Project)
A site for social researchers with an interest in how information and communication technologies are used. Includes case studies and links to ethical issues and technologies pages.
http://responsible-innovation.org.uk/torrii/

Ethical Issues of Emerging Information and Communication Technologies Project (2009-2011)
A project researching ethical issues associated with novel communication technologies. The website links to a variety of the outputs of this project.
ETICA http://www.etica-project.eu/

UK Technology Enhanced Learning Programme
Mobile Ubiquitous and Immersive Technology Enhanced Learning, an Ethical Perspective.
http://www.tlrp.org/docs/MUITEL.pdf

The Internet of Things:
The Pew Centre report on the internet in 2025
http://www.pewinternet.org/2014/05/14/internet-of-things/

ASSOCIATIONS

The Association of Internet Researchers
http://aoir.org/ethics/

E-Research Ethics (Collaboration between Digital Social Research at Oxford University and the Virtual Knowledge Studio for the Humanities and Social Sciences in Amsterdam)
http://eresearch-ethics.org/position/

WEB- BASED MATERIALS ON ETHICS AND DIGITAL DATA

Data Reuse and Licensing Frameworks
http://ands.org.au/publishing/licensing.html

Fair Information Practice Principles
http://en.wikipedia.org/wiki/FTC_Fair_Information_Practice

Marrkula Center for Applied Ethics (at Santa Clara University) Internet Ethics resources:
http://www.scu.edu/ethics/practicing/focusareas/technology/internet/

NatCen Social Research that Works for Society
http://www.natcen.ac.uk/media/282288/p0639-research-using-social-media-report-final-190214.pdf

Social Science Research Ethics pages (Lancaster University)
http://www.lancaster.ac.uk/researchethics/index.html
http://www.lancaster.ac.uk/researchethics/7-1-webres.html

Vison 2Lead.com Social Media Resources
Social Media and Ethical Research Resources
http://vision2lead.com/design/e-research-ethics-resources/

Wellcome Trust: Sharing Research Data to Improve Public Health
http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_peda/documents/web_document/wtvm049648.pdf

GOVERNMENT REGULATIONS ON THE ETHICAL CONDUCT OF RESEARCH

The Economic and Social Research Council Framework for Research Ethics
http://www.esrc.ac.uk/_images/framework-for-research-ethics-09-12_tcm8-4586.pdf

NHMRC 2007 [2014]. National Statement on Ethical Conduct in Human Research
https://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research

National Health and Medical Research Council 2014. Guidelines Under Section 95 of the Privacy Act (1988)
https://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/pr1_guidelines_under_s95_privacy_act_1988_141111.pdf

US Department of Health and Human Services; Secretary's Advisory Committee on Human Research Protections (SACHRP)
http://www.hhs.gov/ohrp/sachrp/index.html

SACHRP: Section on Considerations for Research with the Internet (May 20th 2013, Attachment B)
http://www.hhs.gov/ohrp/sachrp/commsecbytopic/index.html

GUIDELINES

Guidelines on the Ethical Use of Emerging Technologies in Education
Lally, V., Sharples, M., Tracy, F., Bertram, N. & Masters, S. 2012. Muitel an Ethical Perspective (Mobile, Ubiquitous, and Immersive Technology Enhanced Learning)
http://tel.ioe.ac.uk/wp-content/uploads/2010/11/telethics4pp.pdf

Ethical Decision Making and Internet Research Recommendations from the AoIR Ethics Working Committee (Version 2.0 2012)
http://aoir.org/reports/ethics2.pdf

British Educational Research Association
British Educational Research Association (BERA)
Jones, C. (2011). Ethical issues in online research.
https://www.bera.ac.uk/wp-content/uploads/2014/03/Ethical-issues-in-online-research.pdf

British Psychological Association (BPS) Report of the Working Party on Conducting Research on the Internet Guidelines for ethical practice in psychological research online
http://www.bps.org.uk/sites/default/files/documents/conducting_research_on_the_internet-guidelines_for_ethical_practice_in_psychological_research_online.pdf