Research Ethics in Researching Digital Life

This text is an open access version of chapter 4 from the book *Researching Digital Life*.

Please cite as:

Ash, J., Kitchin, R. and Leszczynski, A. (2024) *Researching Digital Life: Orientations, Methods and Practice*. Sage, London.

Overview

This chapter examines:

- the central principles of research ethics;
- ethics frameworks and a situational and reflexive approach to research ethics;
- researching vulnerable communities and sensitive issues, and power relations in research;
- the ethics of using 'found data' produced by others; and
- ethical considerations in producing data using digital media.

Introduction

Research ethics is centrally concerned with ensuring that research is conducted in a manner that adheres to moral principles and societal norms and expectations, and is fair, transparent and non-discriminatory. It recognises that the decisions and practices of conducting research can involve questionable practices that have potentially negative consequences for participants (who are engaged directly in a study), data subjects (those whose data is being used for research without their knowledge) and the communities to which they belong, as well as the researchers themselves, and actively encourages research designs and implementation that seeks to minimise these. Kidder (1981) notes a number of questionable research practices, including:

- involving people in research without their knowledge or consent;
- coercing them to participate;
- withholding information about the true nature of the research;

- otherwise deceiving participants;
- inducing them to commit acts diminishing their self-esteem;
- violating rights of self-determination;
- exposing participants to physical or mental stress;
- invading participants' privacy;
- withholding benefits from some participants;
- not treating participants fairly, with consideration or respect; and
- failing to protect a participant's confidentiality or anonymity.

Four general principles have been central to the development of research ethics (Markham and Buchanan, 2015; Salganik, 2018):

- *respect for people*, in which individuals are treated as autonomous beings who can choose whether to participate, and those with diminished autonomy have additional protections;
- *justice*, in which everyone is treated equally and no group is denied access to, or the benefits from, research;
- *beneficence*, in which all risks and potential harms are minimised and benefits are maximised; and
- *respect for law and public interest*, in which research complies with law, regulations and rules; is transparent and accountable; and aims not to damage communities or undermine public interest and trust.

A related concern is research integrity and the extent to which the research is undertaken in good faith and in line with scientific expectations (e.g., that there is no misconduct such as fraud, fabrication of data or plagiarism).

Formulating appropriate research ethics in research design and application involves a risk/benefit analysis that calculates the probability and potential severity of adverse events, and weighs these up against the potential benefits of the research to the participating community and society in general (Salganik, 2018). In other words, an assessment is made as to whether the research might cause undue stress on participant well-being or cause psychological trauma, or whether the research findings might negatively impact on an individual or community by producing stigma, differential treatment or disinvestment. Such harms might be caused through insensitive or invasive questioning, breaching confidentiality and the disclosure of private information, misuse of data generated (e.g., using information collected for one purpose for a different purpose without the participant's consent) and ill-judged interpretation and communication that might harm a participant's reputation or provides a false impression. The research design of a study needs to anticipate and plan to mitigate against such risks and harms in advance of an empirical study taking place to prevent or limit their occurrence, and to consider possible mitigating procedures in case they do. Ethical uncertainty – where participants and stakeholders have partial information, or are not convinced that suitable protocols and safeguards are in place – can have a chilling effect in terms of securing and maintaining involvement (Salganik, 2018). Practising weak ethics can lead to reputational damage, potential legal action and blacklisting from funding opportunities.

This chapter examines ethics with respect to researching digital life. It starts by detailing key aspects of research ethics in a general sense, discussing ethics frameworks, the concept of contextual integrity, researching vulnerable communities and sensitive issues, power relations in conducting research, the work of institutional review boards and the value of adopting a situational and reflexive approach to ethics. It next considers specific issues related to digital research, divided into two primary sections. First, the ethics of using 'found data' (secondary data available on the internet) are discussed, including privacy, consent, data minimisation, and analysing scraped, hacked and historical data. Second, ethics associated with creating data using digital media are considered, including lurking, covert participation, using commercial crowdsourcing and panel companies, applying data analytics, sharing data and working with the state and businesses.

Key Aspects of Research Ethics

Before considering ethics specifically relating to researching digital life, it is important to be cognisant of broadly applicable issues relating to moral philosophy and ethical practices in conducting research. These include the adoption of an ethical position that guides how ethical decisions in research are made; practising contextual integrity wherein ethical practices are sensitive to context and emerging issues; being aware of particular issues in researching vulnerable communities and sensitive issues, and how power relations within the research process need to be actively managed; considering how researchers themselves also need to be protected from potential risks and harms in conducting research; understanding the role of institutional review boards in overseeing ethical conduct; and a researcher being reflexive and open about their ethical conduct.

Ethics Frameworks

There is no one size fits all, purely instrumental approach to practising research ethics. Indeed, a number of ethical positions and frameworks exist across cultural and disciplinary contexts (Vaughan, 2014). *Deontological* approaches to ethics prioritise action and following agreed-upon rules concerning what is right or wrong over consequences. *Consequentialism* holds that right or wrong should be judged in relation to the consequences of actions, not on whether they comply with rules. *Virtue ethics* places the emphasis on seeking the right thing, rather than on doing and consequences. A feminist *ethics of care* is founded on reciprocity and treating people as one would want to be treated. In each case, ethics is rooted in a different aspect of producing and using knowledge: action, consequence, intent and reciprocity (Vaughan, 2014). Scandinavian countries tend to adopt a deontological approach to research ethics, prioritising the protection of rights for all participants and emphasising dignity, autonomy, equality, and trust (franzke et al., 2020). In contrast, the United States and the United Kingdom are more utilitarian and consequentialist in outlook, willing to consider risking the rights of a few taking part in research for the sake of the greater good. Western ethics tends to focus on individual rights, whereas in non-Western and Indigenous cultures, communal and group rights might be prioritised (franzke et al., 2020). Researchers need to be aware of the prevalent ethical views in the jurisdictions in which their institutions are based and in which fieldwork is to be conducted, and to balance this with their own personal values regarding ethics so far as possible within an institutional review board (IRB) framework (noting that at all times the research needs to be legally compliant). It also means being sensitive to such differences and context when judging the research of others, and being aware that more than one ethically defensible position can be adopted in relation to specific issues (franzke et al., 2020).

Contextual Integrity

Regardless of the ethical framework adopted, in recent years it has been recognised that it is 'impossible to standardize or universalize what constitutes the ethically correct actions in … research contexts, not least because we cannot predict what will happen as a result of our choices' (Markham et al., 2018: 3). Instead, ethicists contend that researchers need to assess carefully, on a case-by-case basis, the specific methods and research design being adopted, and the cultural, regulatory and legal context in which the research is taking place (Hewson, 2016; Lomberg, 2019). In other words, an expectations-based framework for ethical reasoning is applied to consider whether a proposed research design and its possible harms and risks are appropriate within a given context; that is, whether the research design has *contextual integrity*. Contextual integrity is a concept first developed by Helen Nissenbaum (2010) with respect to privacy. She notes that what different communities expect in different circumstances varies, and a one-size-fits-all model of ethical limits calibrated to the highest level of protections that takes no account of context can place unnecessary restrictions on research and curtail valuable studies. As Lomberg (2019: 106) notes, '[t]he principle of contextual integrity invites researchers to dwell on the possible ethical consequences' of their research, and to devise an ethical framing appropriate to the focus, context and vulnerabilities and expectations of participants.

Researching Vulnerable Communities and Sensitive Issues

Much of the ethical concerns pertaining to research relate to protecting the rights of marginalised and vulnerable people, and approaching culturally and politically sensitive issues in an appropriate manner (Markham and Buchanan, 2015). The research ethics literature provides dozens of examples of research projects that have perpetrated deliberate harm on communities in order to observe effects (such as denying essential medical treatment), or else have unintentionally created harm through a lack of planning and foresight (such as causing further mental trauma to victims of abuse through the research design) (Israel and Hay, 2006). It is these studies that have prompted more institutional and regulatory attention being paid to research ethics and the establishment of IRBs by universities. Research in relation to marginalised communities, usually distinguished by social markers such as gender, race, class, disability, sexuality and ethnicity, is often considered sensitive in nature because it potentially has social, political and legal implications, or is considered taboo, sacred or private, or it is actively managed by subjects to limit stigmatisation and negative consequences. Many marginalised communities have a justifiable fear of authority. For example, members of the LGBT community might wish to remain anonymous due to the potential effects of being outed, especially in countries where homosexuality remains illegal, and any research conducted with them must limit any potential threat and protect their identity. Children are a specific class of potentially vulnerable research subjects who have less ability to understand and evaluate participation in research, and weaker autonomy to make decisions regarding consent (Alderson and Morrow 2020). In many jurisdictions, research involving children requires enhanced ethical controls, consent of parents and/or guardians, and police vetting (Monaghan et al., 2013).

Power Relations in Undertaking Research

Related to the issue of researching marginalised communities and sensitive issues is the question of power relations within the research process. Researchers are generally well educated and possess a reasonable degree of social status and cultural capital. They are the ones deciding on a research agenda and formulating a research design. Consequently, there is an asymmetrical power relationship between researchers and those being researched. Such asymmetry can create tensions, particularly if it results in decisions that might cause harm in some way or it alienates research participants where they feel they are not being listened to or are being exploited. England (1994) notes two problematic issues that arise from these power asymmetries: *research tourism* (also referred to as 'academic voyeurism') and *appropriation*. In the case of research tourism, a researcher who does not belong to a social group can act as a voyeur of a marginalised group from a dominant position. The researchers construct the marginalised group's story for them, yet enjoy the privilege of returning to their ordinary life without obligation or responsibility for any consequences of the research that may affect the participants. The marginalised might gain visibility, but not on their terms.

This raises the issue of appropriation; that is, taking a group's knowledge, experiences and skills and materially benefitting from them. Appropriation has been the subject of live debate in fields such as disability studies, queer studies, Black studies, Indigenous studies, decolonial and postcolonial studies, and development studies for several decades. For example, in disability studies, there is a long-standing debate concerning whether research should be conducted on, with or by disabled people, and about the role, actions, motives and consequences of non-disabled researchers undertaking research on disability issues (Kitchin, 2000; Burke and Byrne, 2021). Some have suggested that the traditional 'expert' model of research, wherein non-disabled researchers study disabled people, represents an extractive model of research whereby disabled peoples' knowledge and experiences are appropriated for academic gain (Oliver, 1992). In many cases, this research, however well-meaning, perpetuates the stigmatisation and marginalisation of disabled people. Similar arguments have been made in relation to research and data systems concerning Indigenous communities. As Indigenous scholars have argued, Indigenous communities have experienced centuries of data extraction by non-Indigenous researchers without prior and informed consent, using biased or flawed methodologies, or have been deliberately omitted from official data sources for ends that have rarely benefitted Indigenous peoples (Kukutai and Taylor, 2016; Rainie et al., 2019). Consequently, Indigenous scholars and communities have called for greater data sovereignty; that is, the right to determine and govern how data related to them are generated, analysed, documented, owned, stored, shared and used (Mann and Daly, 2019). In so doing, Indigenous communities are calling for decolonial research methodologies and practices, and for existing data generation systems (such as the production of official statistics) to be decolonised (Walter and Andersen, 2013; Pool, 2016) (see Chapter 2).

Similarly, many social scientists researching communities to which they do not belong favour a research design that seeks to be more inclusive, rebalances power and is sensitive to the concerns of participants. This includes researchers adopting a role that is empathetic and rooted in mutual respect, devolving some aspects of research design and decision-making to research participants, and adopting more participatory approaches in which research is undertaken with and by a community rather than on and about them (England, 1994; Kindon et al., 2007; see Chapter 8). As Elwood and Leszczynski (2018) note, digital scholarship has been relatively slow to consider and address the ways in which it is saturated with and (un)consciously reproduces power.

Protecting Researchers

In addition to protecting research subjects from risks and harms, there is an ethical imperative to do likewise for researchers. In the social media age, undertaking research on certain issues (such as exploitation, discrimination and geopolitics) can provoke strong ideological reactions and expose researchers to online abuse, harassment, doxing (publishing private information about the researcher) and even death threats (franzke et al., 2020). Researching issues such as online hate acts, pornography, criminal activity, terrorism and war can expose researchers to images and first-person accounts that can elicit emotional and psychological reactions and long-term trauma (Roberts, 2019). Principal investigators on projects have a duty of care to their research staff and themselves to consider their safety and psychological well-being, and to put in place suitable procedures to protect staff and deal with any short and long-term issues (franzke et al., 2020).

Institutional Review Boards

A formalised means of assessing the ethics of research is the use of IRBs that aim to ensure compliance with a set of defined acceptable ethical practices. As Hutchinson et al. (2017: 59) note, '[e]thics reviews are a procedural guarantee that normative principles of research integrity have been considered and codified in the research methodology'. Meeting IRB expectations is often the minimum requirement for those researching digital life. The heart of many IRB principles is the FIPPs, developed and adopted by the Organisation for Economic Co-operation and Development (OECD) in 1980. These principles have subsequently underpinned privacy legislation and data protection measures in OECD countries, including the General Data Protection Regulation (GDPR) in Europe. All researchers in the OECD are expected to comply with FIPPs, which set out eight principles of good practice concerning the generation, use, disclosure and sharing of personal data (see Table 4.1), as well as the obligations of data controllers (those who determine the purposes for and the manner in which any personal data are processed) and data processors (those who hold or process data given to them by the data controller) (Solove, 2013). A researcher generating and storing data is a data controller; one who is using secondary or tertiary data is a data processor. Each role has obligations and responsibilities with respect to data subjects and the law.

IRBs can be tricky for social scientists to negotiate for two reasons. First, their foundational principles can often be rooted in medical and health ethics, which are then mapped onto social research with little adaptation or flexibility. Framed around malpractice and potential litigation, bioethics are not well suited to deal with situations 'where normative ethical strategies are unworkable and may threaten academic freedom or participant rights' (Hutchinson et al., 2017: 63). For example, it is unreasonable to seek consent for researching the online activities of a terrorist organisation such as ISIS, or to expect the research strategy or results to be shared with them (Hutchinson et al., 2017). Second, many IRBs fail to acknowledge the variety of ethical frameworks that can be adopted, or permit contextual integrity, being overly rigid and cautious (franzke et al., 2020; Monaghan et al., 2013). One-size fits all assessments can be unhelpful, even if they are well intentioned. Consequently, many social scientists find themselves negotiating with IRBs to persuade them that not only have ethical issues been considered, but the approach being proposed is the most appropriate.

Table 4.1: Fair Information Practice Principles

General principle	General description	Original OECD principle and description
Notice	Individuals are informed that data are being generated and the purpose to which the data will be put.	*Purpose Specification Principle*. The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfilment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.
Choice	Individuals have the choice to opt-in or opt-out as to whether and how their data will be used or disclosed.	*Openness Principle.* There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and

		the main purposes of their use, as well as the identity and usual residence of the data controller.
Consent	Data are only generated and disclosed with the consent of individuals.	*Collection Limitation Principle*. There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.
Security	Data are protected from loss, misuse, unauthorised access, disclosure, alteration and destruction.	*Security Safeguards Principle*. Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data.
Integrity	Data are reliable, accurate, complete and current.	*Data Quality Principle*. Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.
Access	Individuals can access, check and verify data about themselves.	*Individual Participation Principle*. An individual should have the right: 1. to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to her/him; 2. to have communicated to her/him, data relating to her/him within a reasonable time; at a charge, if any, that is not excessive; in a reasonable manner; and in a form that is readily intelligible to her/him; 3. to be given reasons if a request made under subparagraphs(a) and (b) is denied, and to be able to challenge such denial; and 4. to challenge data relating to her/him and, if the challenge is successful to have the data erased, rectified, completed or amended.
Use	Data are only used for the purpose for which they are generated and individuals are informed of each change of purpose.	*Use Limitation Principle*. Personal data should not be disclosed, made available or otherwise used for purposes other than those specified within the notice, except with the consent of the data subject, or by the authority of law.
Accountability	The data holder is accountable for ensuring the above principles and has mechanisms in place to assure compliance.	*Accountability Principle.* A data controller should be accountable for complying with measures which give effect to the principles stated above.

Sources: compiled from OECD (1980) and Minelli et al. (2013).

Situational Ethics and Reflexivity

Research ethics is often practised as a set of compliance rules: a set of procedures and practices that are followed in order to meet expected professional conduct demanded by IRBs. For some they are seen as a nuisance to be adhered to rather than as the minimum, basic moral standards concerning how research is conducted. This is particularly the case in research that frames itself as being scientific, objective, detached, value-free and impartial. For others, research ethics as delineated by IRBs do not go far enough in addressing the politics and power operating within and arising from research. Instead, they advocate for a thorough consideration of ethics and politics at all stages of a project and for the practice of *situational ethics* (Markham et al., 2018). Such an approach involves continually questioning the politics of the research design (in terms of what might be silenced, excluded or privileged through the choices made), the validity of analysis and interpretation, and whose agenda it might be serving. Such views are often grounded in feminist critiques of science that questions its supposed objectivity, neutrality and representativeness.

Donna Haraway (1988, 1991) criticises the dominant epistemology of science for employing what she terms a 'god trick'; that is, claiming to measure and understand the world in a disembodied, emotionless, apolitical view from nowhere that applies to everywhere. Somehow, the researcher is detached from the context and processes of research, the instruments used are in no way reflective of the values of their creators and serve no purpose other than generating objective 'raw' data, and the knowledge produced are representations of reality that are valid outside of cultural interpretation (Propen, 2009). The power of the 'god trick' is that it 'denies the partiality of the knower, erases subjectivities and ignores the power relations involved in all forms of knowledge production' (Kwan, 2007: 24). Moreover, science practised in this way typically privileges those in power and reproduces hegemonic relations and status quo, subjugating, silencing or erasing other perspectives. Instead, Haraway (1988) advocates for an epistemology of partial perspective that she terms *situated knowledges* (see Chapter 2), which calls for transparency and reflexivity in the framing and execution of research projects.

Reflexivity is a key element of a situated approach to knowledge production. To be reflexive is to be self-aware of the politics of choices and decisions being taken, to understand them in context and to consider what the implications of them might be (England, 1994; Rose, 1997). A researcher reflects deeply on epistemology and methodology assumptions, how a project is operationalised, its ethical aspects and consequences for knowledge produced, their positionality, and situated action and interpretation. Reflexivity also involves researchers considering their relational position within the field sites in which they are working (including digital media and platforms), their interactions and asymmetric power relationships with those being researched, how these inflect and mediate the research process, and how the research might affect participants (e.g., negative impacts on their everyday life through the research process itself; or downstream through the introduction of new policies, programmes, regulations, etc.) (Whitson, 2016). In other words, reflexivity involves 'self-critical introspection' and 'self-conscious analytical scrutiny of the self as researcher' *and* the research endeavour (England, 1994: 244), and it recognises that researchers do not 'parachute into the field with empty heads' (England, 1994: 248) but rather arrive with learned knowledge, values, opinions, beliefs, assumptions and 'feelings, failings, and moods' (Stanley and Wise, 1993: 157).

Ethics in Researching Digital Life

Ethical considerations for researching digital life mirror those of non-digital research, but also have a number of specificities and novel aspects (Markham and Buchanan, 2015; Tiidenberg, 2017). For example, given that the internet is scaled globally, which national or international ethical standards should apply to data or subjects that are transnational in character? What are the obligations and legal requirements relating to privacy and data re-purposing for subjects who are engaging in 'public' activities such as posting on social media? How should subjects who are anonymous or may be posing as someone else be treated? How should research on minors be conducted via digital media, and how should consent for participation be sought and verified? What is the legal and ethical status of scraped or hacked data? Do app developers and the owners of platforms have the right to conduct mass experiments on the users of their systems? Moreover, the digital realm raises the question, what counts as a human? Should an avatar in a virtual world or game qualify for ethical protections? Should a robot or a machine that displays some level of autonomy in decision-making be afforded some moral rights (Gunkel, 2018)? In complex systems, such as IoT deployments that might include a number of technologies and stakeholders bound together in complex technical, economic and legal relations, where do responsibilities and consent lie? Similarly, in multi-participant data science (see Chapter 11), who is responsible and accountable for practices and outputs (Leonelli, 2016)? The remainder of the section divides the discussion into two related sections, using a distinction between *found data* (exhaust and secondary data accessible online generated and held by others) and *made data* (produced directly by a researcher and research subjects) (Jensen, 2012). A number of ethical issues and dilemmas are detailed, some of which (such as privacy, consent and working with companies) span both digital and non-digital research.

Ethics and 'Found Data'

Vast quantities of information relating to people, organisations, businesses and their activities are stored as digital data online, which are a potential source of evidence for research. People leave all kinds of data footprints (produced by them) and data shadows (captured by others) through interactions with digital systems, including traces and records of their activities, purchases, location and movement (Kitchin, 2014a). Moreover, they share an enormous amount of information online that in previous generations might have been shared with only a handful of people (family, close friends, employers): CVs or résumés, personal and family stories, family photographs and videos, and personal preferences and thoughts. Digital devices and apps are designed to continuously generate fine-grained data, and digital platforms and infrastructures are *de facto* mass surveillance systems (Zuboff, 2019). These data comprise 'found data': data that are generated, not for the purposes of academic research, but rather for operational, regulatory and optimisation reasons by public, non-profit, business, institutional and other private entities. For instance, users shopping online and posting on social media are engaged in business transactions and social interactions, and do not anticipate their data being repurposed for academic studies.

Found big data are highly relational, being indexical (uniquely tagged to a person, object, location or transaction) and exhaustive (data is generated for all entities in a system rather than being sampled) in nature. What this means is that big datasets contain extensive, fine-grained information about people's digital and digitally mediated interactions with systems, platforms, algorithms, apps, and interfaces. For example, a social media company has information on all posts, shares, likes, and social graphs for all its users; a supermarket chain has detailed time-series purchase records for customers across all its stores; and a telecoms provider has records of all customer activity across its networks. Such records contain a wealth of sensitive information about consumption, interactions, social connections, and beliefs and values. There are clearly ethical issues regarding access to, as well as the handling, analysis, and sharing of, such 'found data'. These include concerns relating to consent, data minimisation, privacy and confidentiality, and uses of scraped or hacked data (see below).

Privacy and Confidentiality

Given the extensive, fine-grained, personal and relational nature of big data, there are clearly issues involving privacy and confidentiality related to how much of the data are publicly accessible or open to scraping, hacking, leaking and analysis. When used in research, found data carry the potential of creating privacy harms (e.g., revealing confidential information) by extending the insights that can be extracted from them through exposure, linkage and analysis, even where the data are publicly available. For example, 'publishing verbatim quotes from a public online discussion forum could lead to them being traced back to source, viewed in context, and individual authors being identified, posing a serious potential threat to participant confidentiality' (Hewson, 2016: 213). High-profile cases that breach privacy include AOL releasing for research more than 20 million search queries from 658,000 users that could be re-identified; a Danish researcher who published the data for 70,000 OKCupid users (a dating site); and 173 million New York taxi journeys being released as open data that were not fully de-identified (Metcalf and Crawford, 2016; Tiidenberg, 2017; Zimmer, 2018). A means to try to mitigate against such harms is to fully de-identify the data and to implement a data management plan (see Chapter 3) that takes data protection and security seriously. De-identifcation can be achieved through several techniques, including anonymisation using pseudonyms, aggregating data, removing selected fields, reducing precision through generalisation, or adding 'noise' (false data) to remove or mask identities (Green et al., 2017). Even when data are de-identified, there may be group privacy issues that need consideration. Group privacy refers to protecting individuals within communities of shared characteristics (such as ethnicity, religion, class, gender, age, health condition, location or occupation) from profiling and differential treatment based on their membership of a group (Rainie et al., 2019; Taylor et al., 2017). Additional means of safeguarding against privacy harms include restricting access to people who can be trusted with data (e.g., people who have undergone ethical training); and storing data on computers with appropriate physical (e.g., locked room) and software (e.g., password protection, encryption) protections (Salganik, 2018).

Consent and Data Minimisation

A central concern with respect to 'found data' is consent, a cornerstone of FIPPs. *Consent* involves securing research subjects' direct permission to participate in a study, and to analyse and share generated data and analysis (Solove, 2013). Most big datasets have not been generated for the purposes of research, and consent to re-use the data has not been sought. Consequently, if their data are re-used, data subjects remain unaware and have no ability to object or withdraw their data from re-use. This was the case in relation to the AOL, OKCupid and New York taxi data mentioned above. In the case of social media or other platform data, some will contend that the data are publicly accessible and therefore are in the public domain and thus open to re-use (Markham and Buchanan, 2015). Here, it is often assumed that any publicly viewable data or data accessible via an API *de facto* has user consent. In such views, there is also a tendency to treat the data as being independent from their subjects, and since humans are not directly involved in the study to believe that the ethical concerns of the research are diminished (Markham et al., 2018). For example, it is not uncommon for data scientists to argue that their research does not require ethical review as it does not involve direct interaction with human subjects (Salganik, 2018).

However, the data are publicly viewable on platforms because their display are essential for their operation; they are not open data, posted so that they can be freely shared and re-used. Moreover, the individuals to which the data refers remain 'data subjects' who have data protection and privacy rights (Buchanan, 2017) and re-use also breaks the data minimisation principle of FIPPs and legislation such as the European Union's GDPR. Data minimisation stipulates that data controllers and processors should only generate data necessary to perform a particular task, that the data are retained for only as long as they are required to perform that task (or as long as legal requirements dictate), and that the data generated should be used only for this task (Tene and Polonetsky, 2012). That is, data should not be re-purposed without consent. Re-using found data that has no research mandate clearly breaks the data minimisation principle. Seeking consent in practice is difficult to achieve given that the data might refer to hundreds of thousands of people and tracking them all down to seek permission is generally impossible. One way that researchers seek to skirt around consent and data minimisation issues is to de-identify the data so that it cannot be traced back to specific individuals. This often involves creating new derived data through generalisation techniques (such as categorisation) that are not subject to the data minimisation principle. This is a less than perfect solution and the ethics of using found data without consent remains a live issue.

Ethics of Using Scraped or Hacked Data

Platforms are commercial enterprises that generate income through the monetisation of their data. While some platforms provide researchers with access to selected data via their APIs (such as TikTok, Meta or YouTube with certain restrictions; Lurie, 2023), most do not openly share their data, although some may enter into specific data-sharing contracts with researchers. However, given that data are publicly accessible as an essential feature of many platforms, they are viewable and amenable to capture through scraping, where a bespoke piece of software (an API) is used to automatically capture data from a platform. An example of scraped data used for research is that collected by Inside Airbnb, which has produced a longitudinal database of scraped short-term rental data from Airbnb for cities across the globe (Scassa, 2019). Data published on government sites are also open to scraping and use in research. For example, Brown (2020) scraped data from the Irish Refugee Appeals Tribunal Archive to assess the practices used by the Irish state to determine asylum in Ireland. The key question with respect to scraping is its legal status. As Scassa (2019) details, while the use of scraped data research often has public interest value, there are uncertainties regarding their ownership, sharing, derivation, intellectual property rights and rights to control data use, as well as how the law views scraping as an activity (e.g., as trespass or theft). Scraped data lacks consent and breaks the data minimisation principle. In using scraped data, or conducting their own scraping, researchers should be mindful of these issues and their possible consequences in terms of legal challenge. Similar concerns relate to the use of hacked data, where non-publicly accessible data have been illegally accessed, copied and shared (Poor, 2017). Such data can be of enormous public value, such as the Snowdon files, Wikileaks and Panama Papers. There are clearly ethical questions in using hacked data in research projects, even when they shed important light on the illegal activities of others. To a large degree, the use of such data is a personal moral decision, though IRBs might seek to block or put limits on such research.

Ethics Relating to Historical Data

Data archives and infrastructures hold vast quantities of historical data, much of it digitised from analogue formats. Much of these data are quite old, episodic and patchy in content, and the individuals to whom the data refer are no longer alive, so issues of consent and privacy dissipate. In countries such as the United Kingdom and Ireland, government data relating to individuals are kept confidential for 100 years before being transferred to national archives, and data relating to government decision-making during key events might be kept confidential for thirty to fifty years. Digital scans of these documents will likewise be kept confidential for the same period. Recent historical data, particularly born digital data, are by comparison more systematic, granular, and indexical; are easier to search and cross-reference; and make it easier for individuals to be found and viewed (Lomberg, 2019). Moreover, whereas ordinary people are less likely to be included in an identifiable form in older historical data sets, they are well represented in big data, such as data generated via social media platforms, even where these data may have been posted by users many years earlier. In addition, what may have previously been considered an unnoticed, niche, private, safe space on the fringes of the web can persist, over time becoming part of the public record in ways unanticipated by former users (Lomberg, 2019). Further, the temporal distance with the present is small as it concerns recent historical data (e.g., a ten-year-old post to a web forum) and any issues may potentially still be live (e.g., a debate and event such as Brexit might unfold over several years and older posts can be easily resurfaced to defend positions or re-ignite flashpoint issues). Consequently, historical data from the recent past containing personally identifiable data need to be subject to ethical practice in similar ways to contemporary data since the same potential harms persist; that is, the principles of respect for people, law and public interest, beneficence and justice still hold. Given time lapses, however, it may be difficult to trace individuals to gain consent, particularly if individuals are no longer active or have left platforms (e.g., deactivated their accounts).

Ethics and 'Made Data'

In contrast to research utilising found data are studies that generate their own data (also known as 'made data'). Here, researchers use prescribed methods to create data. In social science projects, this may involve interactions with people in some fashion; for example, through observation, interviews, surveys and focus groups. In the case of researching digital life, these interactions may be digitally mediated and occur at a distance rather than in person. For example, ethnographies and interviews can be conducted via digital platforms and mobile devices, including messaging apps, email, video conferencing software, online forums and social media (see Chapter 5). While the ethics of such research often mirrors traditional methodologies, this digital mediation does produce some unique challenges, which are discussed below.

Consent

Informed consent is required in all cases where research subjects are knowingly taking part in a research study, and participants must be able to withdraw at any time. In many cases, this is easily dealt with by sending respondents an information sheet and a consent form in advance of a video or email interview, or as the first stage in an app survey. However, the non-proximate nature of the research means that ensuring and verifying that participants have read and understood the information and the nature of their consent, and that they are eligible to take part in terms of age and qualifying criteria (especially when participants are anonymous), is trickier (Hewson, 2016). It is also more difficult to implement effective withdrawal and debriefing procedures, especially when the data generation takes place without researcher presence and involves anonymous subjects with no means of tracing or contacting them (Hewson 2016). In such scenarios, care must be taken to make consent as effective and meaningful as possible; for example, by using double confirmation (re-confirming the initial consent) or continuous consent (seeking consent at each stage of the research process – pre-fieldwork, fieldwork and post-fieldwork phases) and making sure that information is publicly accessible beyond the interaction (e.g., available on a website) (Klykken, 2022). Continuous consent might have particular salience in studies that involve vulnerable populations and sensitive topics. For example, in her study of sexy selfies on social media, Tiidenberg (2018) re-sought consent each time the research shifted phases to ensure that participants continued to be comfortable with the work being undertaken.

Ethics of Lurking, Covert Participation and Researching Closed Groups

Consent is a particular issue in projects where a primary means of generating data is to observe an online community or platform. In general, this takes the form of lurking in which a researcher is virtually present but does not interact with other participants, simply observing and recording activity. Lurking is seen as an attractive method because it observes naturalistic, everyday online interactions, with a reduced risk of the participants changing their behaviour because they know they are being studied (Hine, 2000). This presents a number of related ethical issues. Online communities and platforms can be populated with tens of thousands of participants, many of whom may be anonymous and transitory users, making obtaining individual consent all but impossible. One route is to seek consent from the administrators or owners of a site, group or medium (e.g., a forum or a Facebook page) and to announce publicly that research is underway. However, individual consent will still be absent, and an announcement may not be seen or understood by all. Another approach is to observe covertly, not seeking to gain any form of consent. This, however, challenges the ethical guidelines of many IRBs (Grincheva, 2017). Yet there may be good reasons for using such an approach; for example, the activity is of public interest but is transgressive or illegal, and the research would be impossible to conduct if it were announced; or gaining consent and undertaking the research might fundamentally influence and transform what is being studied. To gain approval from an IRB, researchers must provide strong justification for the covert nature of the study. Similarly, researching a closed group, such as an online support group where membership is by approval only, requires careful handling. The group is closed to protect privacy, and many closed groups are safe havens where members can share their experiences and feelings without being judged (Hård af Segerstad et al., 2017). This may be at odds with the goals of much research, which endeavours to throw light on and examine issues. If a community feels its space, values and trust have been compromised, then significant harm has been inflicted by a study. Care is required in how such groups are approached and consent sought, and how the research data are analysed and disseminated.

Ethics of Using Commercial Crowdsourcing and Panel Companies

Sourcing participants to take part in research studies can be an arduous task. Many academic and commercial researchers have sought to lessen this burden by using third-party platforms and companies to recruit subjects, who then perform the research task online. There are a number of platforms for crowdsourcing research participants, including Mechanical Turk, Prolific Academic, ClickWorker and CrowdFlower (Pittman and Sheehan, 2017). Members of these platforms are offered the opportunity to take part in a study for a fee, usually set per task or unit of time. Crowdsourced participants generally have little screening or sampling controls and are recruited from a general pool. In contrast, research panel companies provide pre-screened, demographically balanced panels of participants who match *a priori* sample criteria and who have usually also agreed to participate in future research, thus providing participant continuity. Panel companies are usually more expensive to employ. Ethical issues relating to the use of crowdsourced platforms and research panels include: fair payment for labour, given that workers receive relatively low wages (often below minimum-wage rates); whether payment enables undue inducement and coercion, encouraging participants to disclose information that they would not otherwise; ensuring effective consent and that principles such as participants can exit at any time are met; and a lack of transparency, since the identity of the researchers commissioning third-party platforms are generally withheld from participants, meaning they cannot evaluate their reputation or trustworthiness (Pittman and Sheehan, 2017). There is live debate concerning whether participants should be paid; the effects of payment on the quality and integrity of research conducted; and issues such as mutual respect, trust and accountability (Grady, 2019; Head, 2009). Some argue that interested volunteers are a better source of research subjects than paid participants, who may feel exploited and care little for study outcomes. Others would favour fair payment to address this issue (Pittman and Sheehan, 2017).

Ethics of Using Data Analytics

Immense computational power and new analytical techniques are now available to merge datasets from disparate sources together, and to sort and sift through datasets and identify patterns and relationships (see Chapter 12). This raises ethical questions concerning the use of data analytics to reveal relations that would otherwise remain hidden, and to act on the findings in ways that might cause harm. In cases where the research is using algorithms and machine learning, there are specific concerns relating to, on the one hand, issues of bias within the learning methods and analytic methods and dependence on third-party datasets (see Chapter 6), and on the other, transparency, review and accountability (Bechmann and Zevenbergen, 2020; Pasquale, 2015). The mobilisation of data analytics in research may produce a false and unfair view of particular populations through a flawed and black-boxed methodology. This might be exacerbated by a reliance on algorithms to interpret and act on data, with an absence of contextual meaning making by domain experts (Markham et al., 2018). There are also questions concerning how the models produced might be used, whether they could be deployed maliciously, and to what extent researchers are responsible for the downstream effects of models if used inappropriately (Bechmann and Zevenbergen, 2020). In such cases, it is incumbent on researchers to consider how their work might be made open, reproducible (when re-analysed using the same methods, the same data will produce the same results) and replicable (when the same methods are applied to new data the same results are produced demonstrating the original study is valid and reliable), in part to help allay ethical concerns (NASEM, 2019).

Ethics of Sharing Data

A predominate trend in academia is for research data to be archived and shared for re-use to facilitate the extraction of additional value and insights. While this is a commendable goal, there are potential ethical issues in sharing research data relating to privacy, confidentiality, data minimisation and potential harms arising from re-use (Corti et al., 2014). In the first instance, care must be taken to ensure that consent covers re-use, particularly as it can be tricky to track down participants later to seek new terms and conditions regarding their data. If data re-use has been consented to, then it is important to ensure that the shared dataset is compliant with privacy and data protection legislation, such as GDPR. This means ensuring that the data are fully de-identified. Qualitative data, such as interview transcripts, can be trickier to de-identify as several elements of non-personal data in the narrative can often be linked to re-identify the interviewee. Solutions include redacting information or converting material into more vague descriptors (Corti et al., 2014). If the sound file is also being deposited, then personal identifiable material should be bleeped out and the voice can be disguised by altering the pitch and tone; similarly, faces in photos or videos can be pixilated. It may also be necessary to implement tiered access (e.g., limiting access to authorised users); embargos (e.g., data are only released after a set period of time); or stipulations that the data only be used for particular purposes, such as replication and reproducibility, but not for others, such as being enrolled into commercial databases and monetised in some way. Implementing all these techniques can be time consuming and expensive.

Ethics of Working with States and Companies

Much research concerning digital life focuses on the work of states and businesses and how they use digital technologies to regulate and manage society and to produce profit. This research is often critical in nature, detailing how these entities' actions (re)produce structural relations, and questioning their logics and practices. Unsurprisingly, then, there are concerns about academics working with such actors. Indeed, there is a long-running debate in the social sciences concerning the independence and purpose of academic research (see Fuller and Kitchin, 2004). On one side are those who believe that academic research should necessarily maintain a separation from state and industry to ensure critical distance and scientific autonomy, and to avoid being co-opted into and legitimising state and industry actions (Allen, 2011). On the other side are those who believe that academia should work with state and industry to tackle societal and fundamental problems by pooling knowledge, expertise and resources (Bastow et al., 2014). It is a personal choice to work with state or business partners, but if one does choose to collaborate with these entities, there are a number of issues to keep in mind. Agreements concerning ethics might operate solely at the level of compliance with legal requirements, as achieving an alignment with respect to moral philosophy might be difficult and partners might have quite different ambitions regarding research outcomes and how they will be applied in practice. Moreover, industry partners might not operate in alignment with the IRB strictures demanded by universities or funding agencies, and lack independent oversight mechanisms (Hoffman and Jonas, 2017). They may also demand the signing of non-disclosure agreements that prevent academic researchers from being fully transparent and from voicing their concerns about aspects of a study. However, it might well be that the only way to examine the platforms and services of digital technology companies is to enter into a working arrangement with them, meaning that a compromise of usual ethical standards enables important access to hidden assets, practices and outcome. While potentially of benefit, a danger here is the partnership enacts a form of ethics washing (Wagner, B., 2018), with the university collaboration legitimatising problematic practices.

Summary

- This chapter has detailed the various ethical issues that need to be negotiated in researching digital life. In particular, it has examined specific ethical issues related to using ‘found’ and ‘made’ data.
- A number of general issues, such as privacy, consent, data minimisation and working with state and business partners, need to be considered in designing and implementing a project, as well as issues that are more specific to researching digital life, such as using scraped or hacked data, lurking or crowdsourcing.
- All research is saturated with ethics and politics in its formulation, execution and outcomes and careful attention must be paid to the power relations at play within a project and the various harms that undertaking research might have on participants and communities. This is particularly the case for researching vulnerable communities and sensitive issues.
- Researchers might seek to ensure that their research practices are conducted in a neutral, detached, objective manner that minimise harms to people and systems under investigation, but striving for such a goal involves active, reflexive attention to ethics throughout the *lifetime* of a project as circumstances change.
- At a minimum, a project should comply with IRB principles. Better still, a thoroughly moral, reflexive and situated approach to conducting research and ethics should be adopted that aims to minimise detrimental practices and harmful effects to participants while producing insightful and useful knowledge.

Recommended Reading

- franzke, a.s., Bechmann, A., Zimmer, M., Ess, C., and the Association of Internet Researchers (2020) *Internet Research: Ethical Guidelines 3.0.* https://aoir.org/reports/ethics3.pdf (accessed 16 August 2023).

  This report details a comprehensive set of ethical issues in conducting digital research and how to approach them.
- Markham, A. N., and Buchanan, E. (2015) ‘Ethical considerations in digital research contexts’, in Wright, J.D. (ed.), *Encyclopedia for Social & Behavioral Sciences*. Elsevier, Waltham, MA, pp. 606–13.

  A clear, concise introduction to research ethics in internet and big data research.

- Utrecht Data School (n.d.) *Data Ethics Decision Aid for Researchers,* https://deda.dataschool.nl/en/. (accessed 5 Sept 2023).
  A useful tool for prompting critical thought on the ethics of a project.
- Zimmer, M., and Kinder-Kurlanda, K. (eds) (2017b) *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts*. Peter Lang, New York.
  Provides a wide-ranging discussion of ethics in conducting research using social media platforms and digital methods, including detailed case studies.
- Zimmer, M. (2018) 'Addressing conceptual gaps in big data research ethics: An application of contextual integrity', *Social Media + Society*, 4(3): 1–11.
  Provides a useful heuristic for thinking through the potential ethical issues of big data research.

References

Alderson, P. and Morrow, V. (2020) *The Ethics of Research with Children and Young People: A Practical Handbook*, 2nd edn. Sage, London.

Allen, C. (2011) 'Against dialogue: Why being critical means taking sides rather than learning how to play the "policy research" game', *Dialogues in Human Geography*, 1(2): 223–7.

Bastow, S., Dunleavy, P., and Tinkler, J. (2014) *The Impact of the Social Sciences: How Academics and Their Research Make a Difference*. Sage, London.

Bechmann, A., and Zevenbergen, B. (2020) *AI and Machine Learning: Internet Research Ethics Guidelines, IRE 3.0 Companion 6.1.* Association of Internet Researchers, https://aoir.org/reports/ethics3.pdf (last accessed 12 June 2023)

Brown, S. (2021) *Evidence and Absence in the Archives: A Study of the Irish Refugee Appeals Tribunal Archive to Assess the State Practice of Determining Asylum in Ireland*. PhD thesis, Maynooth University, Maynooth.

Buchanan, E. (2017) 'Considering the ethics of big data research: A case of Twitter and ISIS/ISIL', *PLoS ONE*, 12(12): e0187155.

Burke, C., and Byrne, B. (eds) (2021) *Social Research and Disability: Developing Inclusive Research Spaces for Disabled Researchers*. Routledge, Abingdon.

Corti, L., van den Eynden, V., Bishop, L., and Woollard, M. (2014) *Managing and Sharing Research Data: A Guide to Good Practice*. Sage, London.

Elwood, S., and Leszczynski, A. (2018) ‘Feminist digital geographies’, *Gender, Place & Culture*, 25(5): 629–44.

Fuller, D., and Kitchin, R. (eds) (2004) *Radical Theory, Critical Praxis: Making a Difference Beyond the Academy?* Praxis E-Press. kitchin.org/wp-content/uploads/2020/05/RTCP_Whole.pdf

England, K. (1994) ‘Getting personal: Reflexivity, positionality, and feminist research’, *The Professional Geographer*, 46(1): 80–9.

franzke, a.s., Bechmann, A., Zimmer, M., Ess, C., and the Association of Internet Researchers (2020) *Internet Research: Ethical Guidelines 3.0*. https://aoir.org/reports/ethics3.pdf (accessed 16 August 2023).

Grady, C. (2019) ‘The continued complexities of paying research participants’, *The American Journal of Bioethics*, 19(9): 5–7.

Green, B., Cunningham, G., Ekblaw, A., Kominers, P., Linzer, A., and Crawford, S. (2017) *Open Data Privacy*, The Berkman Klein Center for Internet & Society Research Publication, 2017–1. Harvard University, Cambridge, MA. https://ssrn.com/abstract=2924751 (accessed 22 August 2023).

Grincheva, N. (2017) ‘Museum ethnography in the digital age’, in Zimmer, M., and Kinder-Kurlanda, K. (eds) *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts*. Peter Lang, New York, pp. 187–94.

Gunkel, D.J. (2018) *Robot Rights*. MIT Press, Cambridge, MA.

Haraway, D. (1988) ‘Situated knowledges: The science question in feminism and the privileges of partial perspective’, *Feminist Studies*, 14(3): 575–99.

Haraway, D.J. (1991) *Simians, Cyborgs, and Women: The Reinvention of Nature*. Routledge, New York.

Hård af Segerstad, Y., Kasperowski, D., Kullenberg, C., and Howes, C. (2017) ‘Studying closed communities online: Digital methods and ethical considerations beyond informed consent and anonymity’, in Zimmer, M., and Kinder-Kurlanda, K. (eds) *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts*. Peter Lang, New York, pp. 213–25.

Head, E. (2009) ‘The ethics and implications of paying participants in qualitative research’, *International Journal of Social Research Methodology*, 12(4): 335–44.

Hewson, C. (2016) ‘Ethical issues in digital methods research’, in Snee, H., Hine, C., Morey, Y., Roberts, S., and Watson, H. (eds) *Digital Methods for Social Sciences: An*

Interdisciplinary Guide to Research Innovation. Palgrave Macmillan, New York, pp. 206–21.

Hine, C. (2000) *Virtual Ethnography*. London, Sage.

Hoffman, A.L., and Jonas, A. (2017) 'Recasting justice for Internet and online industry research ethics', in Zimmer, M., and Kinder-Kurlanda, K. (eds) *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts*. Peter Lang, New York, pp. 3–18.

Hutchinson, J., Martin, F., and Sinpeng, A. (2017) 'Chasing ISIS: Network power, distributed ethics and responsible social media research', in Zimmer, M., and Kinder-Kurlanda, K. (eds) *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts*. Peter Lang, New York, pp. 57–71.

Israel, M. and Hay, I. (2006) *Research Ethics for Social Scientists*. Sage, London.

Jensen, K.B. (2012) 'Lost, found, and made', in Volkmer, I. (ed.) *The Handbook of Global Media Research*. Wiley-Blackwell, Oxford, pp. 433–50.

Kitchin, R. (2014a) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Sage, London.

Klykken, F.H. (2022) 'Implementing continuous consent in qualitative research', *Qualitative Research*, *22*(5): 795–810.

Kukutai, T., and Taylor, J. (2016) 'Data sovereignty for indigenous peoples: Current practice and future needs', in Kukutai, T., and Taylor, J. (eds), *Indigenous Data Sovereignty: Towards an Agenda.* Australian National University Press, Canberra, pp. 1–22.

Kidder, L.H. (1981) *Research Methods in Social Relations*. Harcourt Brace, London.

Kindon, S., Pain, R., and Kesby, M. (2007) *Participatory Action Research Approaches and Methods*. Routledge, Abingdon.

Kitchin, R. (2000) 'The researched opinions on research: Disabled people and disability research', *Disability and Society,* 15(1): 25–48.

Kwan, M.-P. (2007) 'Affecting geospatial technologies: Toward a feminist politics of emotion', *The Professional Geographer,* 59(1): 22–34.

Leonelli, S. (2016) 'Locating ethics in data science: responsibility and accountability in global and distributed knowledge production systems', *Philosophical Transactions of the Royal Society A*, 374: 20160122.

Lomberg, S, (2019) 'Ethical considerations for Web archives and Web history research', in Brugger, N., and Milligan, I. (eds) *The Sage Handbook of Web History*. Sage, London. pp. 99–111.

Lurie, E. (2023) ‘Comparing platform research API requirements’, *Tech Policy Press*, 22 March, https://techpolicy.press/comparing-platform-research-api-requirements/ (accessed 22 August 2023).

Mann, M., and Daly, A. (2019) ‘(Big) data and the north-in-south: Australia’s informational imperialism and digital colonialism’, *Television & New Media*, 20(4): 379–95.

Markham, A. N., and Buchanan, E. (2015) ‘Ethical considerations in digital research contexts’, in Wright, J.D. (ed.), *Encyclopedia for Social & Behavioral Sciences*. Elsevier, Waltham, MA, pp. 606–13.

Markham, A.N., Tiidenberg, K., and Herman, A. (2018) ‘Ethics as methods: Doing ethics in the era of big data research – Introduction’, *Social Media + Society,* 4(3): 1–9.

Metcalf, J., and Crawford, K. (2016) ‘Where are human subjects in Big Data research? The emerging ethics divide’, *Big Data & Society*, 3(1): 1–14.

Minelli, M., Chambers, M., and Dhiraj, A. (2013) *Big Data, Big Analytics*. John Wiley & Sons, Hoboken, NJ.

Monaghan, L.F., O’Dwyer, M., and Gabe, J. (2013) ‘Seeking university research ethics committee approval: the emotional vicissitudes of a ‘rationalised’ process’, *International Journal of Social Research Methodology*, 16(1): 65–80.

NASEM (National Academies of Sciences, Engineering, and Medicine) (2019) *Reproducibility and Replicability in Science*. The National Academies Press, Washington DC.

Nissenbaum, H. (2010) *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, Stanford, CA.

OECD (1980) *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. OECD, Paris, www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm (accessed 2 December 2020).

Oliver, M. (1992) ‘Changing the social relations of research production’, *Disability and Society*, 7(2): 101–14.

Pasquale, F. (2015) *The Black Box Society*. Harvard University Press, Cambridge, MA.

Pittman, M., and Sheehan, K. (2017) ‘Ethics of using online commercial crowdsourcing sites for academic research: The case of Amazon’s Mechanical Turk’, in Zimmer, M., and Kinder-Kurlanda, K. (eds) *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts*. Peter Lang, New York, pp. 177–86.

Pool, I. (2016) ‘Colonialism’s and postcolonialism’s fellow traveller: The collection, use and misuse of data on indigenous people’, in Kukutai, T., and Taylor, J. (eds), *Indigenous Data Sovereignty: Towards an Agenda*. Australian National University Press, Canberra, pp. 57–76.

Poor, N. (2017). ‘The ethics of using hacked data: Patreon’s data hack and academic data standards.’ in Zimmer, M., and Kinder-Kurlanda, K. (eds) *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts*. Peter Lang, New York, pp. 277–80.

Propen, A. (2009) ‘Cartographic representation and the construction of lived worlds: understanding cartographic practice as embodied knowledge’, in Dodge, M., Kitchin, R., and Perkins, C. (eds) *Rethinking Maps*. Routledge, London, pp. 113–30.

Rainie, S.C., Kukutai, T., Walter, M., Figueroa-Rodríguez, O.L., Walker, J., and Axelsson, P. (2019) ‘Issues in open data: Indigenous data sovereignty’, in Davies, T., Walker, S., Rubinstein, M., and Perini, F. (eds), *The State of Open Data: Histories and Horizons*. African Minds and International Development Research Centre, Cape Town and Ottawa, pp. 300–19.

Roberts, S.T. (2019) *Behind the Screen*. Yale University Press, London.

Rose, G. (1997) ‘Situating knowledges: positionality, reflexivities and other tactics’, *Progress in Human Geography*, 21(3): 305–20.

Salganik, M.J. (2018) *Bit by Bit: Social Research in the Digital Age*. Princeton University Press, Princeton, NJ.

Scassa, T. (2019) ‘Ownership and control over publicly accessible platform data’, *Online Information Review*, 43(6): 986–1002.

Solove, D. (2013) ‘Privacy management and the consent dilemma’, *Harvard Law Review*, 126: 1880–903.

Stanley, L., and Wise, S. (1993) *Breaking Out Again: Feminist Ontology and Epistemology*. Routledge, London.

Taylor, L., Floridi, L., and van der Sloot, B. (eds) (2017) *Group Privacy: New Challenges of Data Technologies*. Springer, Cham.

Tene, O., and Polonetsky, J. (2012) ‘Big data for all: Privacy and user control in the age of analytics’, *Northwestern Journal of Technology and Intellectual Property*, 11(5): 240–73.

Tiidenberg, K. (2018) ‘Ethics in digital research’, in Flick, U. (ed.) *The Sage Handbook of Qualitative Data Collection*. Sage, London, pp. 466–79.

Utrecht Data School (n.d.) *Data Ethics Decision Aid for Researchers,* https://deda.dataschool.nl/en/. (accessed 5 Sept 2023).

Vaughan, L. (2014) *Beginning Ethics: An Introduction to Moral Philosophy*. W.W. Norton, New York.

Wagner, B. (2018) ‘Ethics as an escape from regulation: From ethics-washing to ethicsshopping?’, in Hildebrandt, M. (ed.) *Being Profiling. Cogitas ergo sum*. Amsterdam University Press, Amsterdam, pp. 84–9.

Walter, M., and Andersen, C. (2013) *Indigenous Statistics: A Quantitative Research Methodology*. Left Coast Press, Walnut Creek, CA.

Whitson, R. (2017) ‘Painting pictures of ourselves: Researcher subjectivity in the practice of feminist reflexivity’, *The Professional Geographer*, 69(2): 299–306.

Zimmer, M., and Kinder-Kurlanda, K. (eds) (2017b) *Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts*. Peter Lang, New York.

Zimmer, M. (2018) ‘Addressing conceptual gaps in big data research ethics: An application of contextual integrity’, *Social Media + Society*, 4(3): 1–11.

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. Profile Books, New York.