



Making Ethical Decisions in an Online context: Reflections on using blogs to explore narratives of experience

Helene Snee

Sociology, University of Manchester

Abstract

Internet research methods can present some challenging ethical dilemmas. Although they are subject to the same guiding principles as ‘offline’ research, it can be difficult to apply these online given the blurring of boundaries presented by digitally-mediated environments. This paper considers a study that utilised personal blogs as primary data to outline two common ethical tensions in internet research: whether online communications can be considered public or private, and whether the people who produce them can be considered subjects or authors. The study examined the narratives of young people who took gap years overseas as represented in their travel blogs. While the blogs were technically public, they contained personal information, and individual perceptions or expectations of privacy can be different. The paper also explores how to establish if human subjects are involved in internet research, including the difficulties of protecting identity and the case for recognising authorship when appropriate. In line with contemporary scholarship on internet research ethics, this paper highlights the need for a contextual approach that recognises the specificities of the communications studied, the methods employed to generate and analyse data, and how the research is disseminated. The decisions made in the gap year study are critically evaluated, and alternative options presented, including a focus on ensuring that data are not linked to individuals without consent. The article aims to contribute to dialogue and debate in online research ethics through offering some reflections on the course of action taken.

Keywords: ethics; internet research; blogs; narratives; privacy.

Introduction

This paper reflects on the ethical decisions made during a study that explored the narratives presented in young people’s travel blogs to tell the story of their ‘gap year’ experiences (Snee 2013). Scholars of internet research suggest that the best way to negotiate the complex ethical dilemmas associated with online research is to adopt a ‘bottom-up’¹, contextual approach (Markham & Buchanan 2012). Such a strategy is undoubtedly appropriate as the wide variety of online spaces and interactions, along with the research methods that may be utilised to collect and analyse them, mean that blanket ethical guidelines could not cover all eventualities and scenarios in internet research. The downside, however, is that it can be difficult for researchers to establish the best way forward if they wish to explore social life using the rich data available online. How can we determine if our research is ethical? Internet researchers have offered reviews, guidelines and case studies that enable us to think through how we might conduct online research in an ethical manner. As such, this paper hopes to contribute to these resources by offering some reflections on the ethics of qualitative blog analysis.

This paper does *not* provide the ‘right answers’ about internet research ethics or the ethics of blog analysis more specifically. This would not be suitable in any case, given the importance of a contextual approach outlined above. What could be the best way forward for this research project may not necessarily apply to another. Yet I have found myself reflecting upon, and questioning, the ethical decisions I made at the time. I began the study outlined below as a PhD student and novice researcher, and the paper critically reflects upon the decisions I made to provide insight into debates in online research. To provide some background to this exploration, this paper provides an overview of the gap year project, and the methods used to generate and analyse the blog data. I then outline some of the key issues associated with conducting research online, and use the gap year project to focus on two central concerns: whether the research dealt with human subjects, and whether the research used public or private data. The gap year blogs included in the sample were classified as texts that were personal, but not private (Hookway 2008). However, I conclude by considering the alternative decisions I could have made, and ask: ‘Did I do the right thing?’

The gap year project

The research did not focus on the practice of blogging itself. Instead, it was designed to explore young people’s stories of their gap year experiences. A gap year is typically conceived as a period for a young person to travel independently, volunteer overseas (often in less developed countries), engage in cross-cultural learning, and other such ‘worthwhile’ activities. This is the popular understanding of gap years, although they can be taken at a variety of stages in a person’s life, and do not necessarily involve travel. Taking a gap year between school and university is a popular activity in the UK, and is seen as a legitimate way for young people to have a break before they progress to the next stage of their education. As noted by Simpson (2004), gap years are an institutionalisation of youth travel practices in an industry that receives the support of the UK government and education sector. Research into the phenomenon of gap year travel has questioned its role in reproducing global inequalities, hierarchies of taste, educational inequalities, and in identity work (e.g. Heath 2007, King 2012, Simpson 2004).

The study set out to analyse how young people’s narratives of gap year experiences were framed. The approach taken drew upon Goffman’s concept of ‘frame’, the ‘definitions of a situation [that] are built up in accordance with principles of organization which govern events... and our subjective involvement in them’ (Goffman 1974: 10-11). Frames are how we make sense of experiences, but also how we can represent them. Blogs provide a way of studying the frames that young people used to present their gap year stories. This is a form of unobtrusive research, drawing on data that is not produced by elicitation from subjects, for example physical traces, non-participant observation or documentary sources (Lees 2000; Pauwels 2006). Unpacking some of the different resources that the participants drew upon to understand their practices and communicate them to others gave an insight into how individual and collective understandings guide the stories that are told.

Blog analysis is comparable with documentary analysis, exploring the ‘representational world’ of individuals and groups using material archived online (Arosio 2010). Arosio (2010) highlights the similarities of blogs with diaries, in that they are personal documents produced in real time with no precise addressee. There are also crucial differences, as they are visible to anyone with online access, and they are interactive (2010: 31). In addition, there are usually ‘implicit, if not explicit, audiences’ for a blog (Hookway 2008: 96). In the case of travel blogs, for example, they can be a way for friends and family to follow the blogger’s journey and connect with them while they are away (Germann Molz 2012). One of the key advantages of blog analysis is that the narratives found in personal blogs are spontaneous, and offer something different to those that would be available through interaction with a researcher (Hookway 2008). Spontaneous, in this sense, refers to how

personal documents are produced by people ‘carrying out their activities.. without any link with the research goals’ (Arosio 2010: 25). Of course, this is a particular *version* of gap years. These representations contain what the bloggers think is important to record and communicate.

The blog sample was constructed by searching for the phrase ‘gap year’ on two blog search engines: Google Blog Search and Technorati; and three blog platforms: Myspace, LiveJournal, and the travel journal site Globenotes. The search results generated were examined and blogs were included in the sample if they met the criteria for the study: they were written by young people from the UK who were taking a gap year with some time spent overseas between school and university. The process of data collection continued until a sample of 40 blogs was constructed that was balanced in terms of gender of author(s), and included a broad range of types of gap year travel activities. Once the sample was constructed, the bloggers were contacted for interview, either via a contact email address that the author provided on their blog; via a message function on the blog; or by leaving a comment. The bloggers were asked to contact me at my university email address if they were interested. The rationale for contacting the bloggers for interview was to enable reflection on the themes identified through a review of existing work on gap years and in the course of the blog analysis. It also enabled the young people to talk about the elements that the blogs did not answer, such as their motivations for writing blogs, and what was not included in their online accounts.

Out of the initial sample of 40 blogs: nine bloggers were interviewed; seven expressed an interest but did not respond to further messages or could not commit to an interview; one agreed to be interviewed but did not attend; and 23 did not respond. During these communications, it transpired that one blogger did not meet the sample criteria (having taken her year out on sabbatical half-way through her university course for health reasons). This resulted in a final sample of 39 blogs. In order to manage the large volume of data that was generated, each blog was saved as a text file and imported into Atlas.ti. The files were then coded via an iterative process, and analysed using qualitative thematic analysis, influenced by Kelle’s (1997) description of ‘qualitative induction’ and ‘abductive inference’. This both describes specific phenomenon with reference to existing categories and new and atypical events presented by the data (Kelle 1997).

A range of practical and methodological issues were encountered during the research process, including managing a significant amount of multimedia data and evaluating the ‘authenticity’ of blogs (see Snee, 2012 for a more detailed discussion). Of particular concern was ensuring that the data was dealt with ethically, as this was drawn from personal blogs written by young people. There are also ethical implications of using unsolicited material not produced for the purposes of social research. Moreover, making contact and interacting with the bloggers during the process of asking for interviews problematises the project’s status as unobtrusive research. The following section outlines the field of internet research ethics in order to introduce the key questions that were considered during the gap year study. This concerns whether the research was dealing with human subjects, and whether the blog posts could be considered private.

Debates in internet research ethics

Ethical guidelines

The research was subject to the ‘traditional’ ethical considerations that govern any sociological project, and the guiding principle – of doing no harm to a research subject – remained central. Discussions of internet research ethics have highlighted that this principle needs to be upheld. The Association of Internet Researchers (AoIR) have produced ethical guidelines that provide an overview of the debates and tensions in negotiating online research (Ess et al. 2002; Markham & Buchanan 2012). The latest version acknowledges that: ‘the basic tenets shared by these policies include... respect for persons, justice, and beneficence... We accept them as basic to any research endeavour’ (Markham & Buchanan 2012: 4).

The online environment is a tricky one to negotiate in terms of deciding how to apply the principle to do no harm. Grinyer (2007:1) notes that certain characteristics of internet research pose new dilemmas for researchers that may not be covered by existing ethical guidelines, and points to a relevant extract from the British Sociological Association's Statement of Ethical Practice: 'Eliciting informed consent, negotiating access agreements, assessing the boundaries between the public and the private, and ensuring the security of data transmissions are all problematic in Internet research' (BSA 2002: 5). Similarly, Frankel & Siang (1999) highlight:

...the ability of both researchers and their subjects to assume anonymous or pseudoanonymous [sic] identities online, the complexities of obtaining informed consent, the often exaggerated expectations, if not the illusion, of privacy in cyberspace, and the blurred distinction between public and private domains (Frankel & Siang 1999: 1-2).

Ethical standards for internet research are subject to some debate. The medium itself is still formative and new phenomena – such as blogs – are emerging (Eynon et al. 2008). For example, the first version of the AoIR guidelines were produced in 2002, before the integration of social network sites such as Facebook into everyday life. Previous debates considered whether 'new' ethical procedures were required for online research, or whether the fundamental principles of research ethics should remain unchanged. A case has been made that specific guidelines are required, as this will protect both the public and support the institutions of social science, and therefore be useful to both researchers and wider society (Madge 2006). The mediated research context, and the blurred boundaries described above, have led some to argue for new rules (e.g. Frankel & Siang 1999). Others, however, have argued that the ethical questions are similar to those encountered in other media or offline research (e.g. Walther 2002). Walther argues that highlighting internet research as a particular problem is 'methodological myopia', which means that other methods are seen to be less problematic (cited in Orton-Johnson 2010). Drawing on these two perspectives, Eynon, Fry and Schroeder suggest that while internet research may not be more risky, the risks might be different, particularly as it can be difficult to assess such risks through academic precedents or the reactions of the users studied (2008: 27). Another issue is that existing arrangements may not engage with internet research ethics in an appropriate way. Internet research is singled out in the ESRC framework for research ethics as particularly risky, but Orton-Johnson (2010) argues this homogenises a variety of methods, approaches and practices. This is of particular concern in a contemporary climate where the distinction between online and offline may not be useful (Orton-Johnson 2010: 2.3).

A contextual approach

These debates have led to a general consensus that core ethical standards should be applied to internet research, but that these need to be tailored to the specific research context. This consensus is encapsulated in the AoIR guidelines (Ess et al., 2002, Markham & Buchanan 2012) and put forward by a number of Internet scholars (Eynon et al. 2008, Lomborg 2013). The rationale is that it is not possible to provide universal guidelines because there are an 'almost endless' range of unique situations that could be presented by the umbrella term of 'internet research' (Markham & Buchanan 2012: 7). A pluralist framework thus acknowledges: cultural and national differences (Lomborg 2013: 20); that 'the internet is many things to many people' (Eynon et al. 2008: 26); and the wide range of methods that are employed by internet researchers (Orton-Johnson 2010). Subsequently, the approach advocates that 'ethical judgement in internet research must rely on an inductive, contextual assessment of the specific case in question, and that different ethical choices may equally be legitimate and sound' (Lomborg 2013: 21). This requires an assessment, for example, of the research design, methods and conceptual frameworks employed (Eynon et al. 2008: 23-4). One crucial

exercise is to review previous scholarship for examples of how ethical decisions were negotiated (Markham & Buchanan 2012: 12). The purpose of this article is to contribute to resources for this exercise.

There are some common concerns associated with internet research that need to be considered using this context-based approach. Research governance is based on the idea of protecting human subjects. Yet as will be explored in more detail below, establishing whether human subjects are involved in internet research is not straightforward. A related issue is whether the data being studied involves a person or a text. Finally, we can ask whether there are different understandings of privacy in different online environments. The recent AoIR guidelines critically reflect on the concept of defining human subjects in internet research, and highlight two key tensions: public / private, and data (texts) / persons (Markham & Buchanan 2012). These tensions had to be negotiated in the gap year study, and a decision made as to whether the blogs were public or private, and whether the bloggers were subjects or authors. Resolving these tensions had a significant impact on how the research progressed, and the need to obtain informed consent, protect privacy, and preserve anonymity. I now expand on these issues, and how they were resolved in the gap year study, with reference to a contextual approach to research ethics. As will be explored in the concluding section, I take a critical view of the decisions made, and put forward an alternative way that this particular research context could have been negotiated.

Blogs as public or private?

Changes to public / private boundaries

One of the most frequently discussed issues in internet research ethics is the blurring of boundaries. Of particular concern is how to define what is public and what is private. As noted by Hine (2011), although we might be able to easily access data using unobtrusive methods, this does not make this 'ethically available' (2011: 3). While the conceptual distinction between public and private is not universal, and is relatively new in human history, consensus in the West on what is made public and what is kept private developed around the nineteenth century (Ford 2011: 552). Internet researchers over the past 20 years have attempted to negotiate how online communications have affected this boundary. The nature of contemporary social media complicates this further. Increasingly, personal information is volunteered in blogs and social networking sites. Beer & Burrows (2007) place this in the context of wider shifts in the sense and values of privacy, which can also be seen in other cultural and social trends such as reality television and celebrity culture (Beer & Burrows 2007).

What we actually *mean* by private has changed. As Beer & Burrows note, people now curate their own archives of everyday life on social media, including: 'their preferences, choices, views, gender, physical attributes, geographical location, background, employment and educational history, photographs of them in different places, with different people and different things' (Beer & Burrows, 2007). If someone had wanted to analyse such data in the past, they would need to have had access to the individuals in question, or relied on donated historical archives. Ford (2011) thus sees the social web as key to the reshaping of the public / private distinction. These boundaries are being tested by other developments, such as surveillance technologies and the use of ICTs in work practices, however for Ford the social web has:

...brought the tensions between public and private to the forefront of the collective consciousness...
Internet users seem to be airing much of their 'private' lives in the 'public' realm of the internet'
(Ford 2011: 557).

Facebook is a notable example of how people are 'comfortable with far more transparency than had been anticipated' (Miller 2011: 173). For Miller (2011), Facebook is changing our relationship to privacy. However, not everyone has the same ideas about what should be shared. Moreover, Miller suggests we should

view the ‘public’ space of Facebook as ‘an aggregate of private spheres’ (2011: 174). Wider shifts between public and private can also be understood using Berlant’s definition of intimate publics, ‘constituted by strangers who share common texts and things’ (2008: viii). Media texts, and the emotional connections people make to them, provide a collective experience which engender a sense of belonging. For example, Morrison (2011) argues that ‘personal mommy blogging’, in which mothers share stories of parenting, creates shared communities of emotional responses. However, she suggests that these online practices are different to the consumption of mass media texts that Berlant describes, as they are small, reciprocal networks of ‘interlinked individual sites of textual production’ (Morrison 2011: 51).

Thus, while the boundary between public and private has blurred, we can question whether the availability of information on the internet necessarily makes this information ‘fair game’. Interventions in the ‘mommy blogging’ networks could encroach on a community that provides emotional support to its members. Heath et al (1999) suggest that ‘lurking’, in which researchers covertly collect data online, encroaches privacy and creates an unequal power relationship (cited in Grinyer 2007: 2). In essence, the researcher knows all about the person producing the information found online, but this person knows nothing about the researcher. Moreover, online communications obtained unobtrusively are, by definition, not produced for researchers. Considering ‘online diaries’ from a biographical perspective, McNeill (2003) notes her own sense of unease in reading these forms of communication. Although such diaries are produced for readers both ‘imagined and actual’, she felt some alienation in consuming these narratives as she felt that she was *not* the intended audience (McNeill 2003: 32).

Applying this to the gap year study, the young bloggers may have been producing their stories for their friends and family, or even the wider audience – but they were not produced for a researcher to analyse. Conversely, we might argue that if something is put into the public domain, control over what happens to this data is relinquished. Deciding how to proceed was an important issue as while anyone with an internet connection could access the blogs included in the gap year sample, they often contained the young person’s full name, photographs of them and their friends and their email address, alongside some intimate reflections. A study of teenagers’ social media presence found that blogs and social network sites revealed a ‘considerable amount of personal information’ (Huffaker 2006: 6). It was necessary to consider whose the responsibility it was to protect this information, even though it was available online, and to look for ways to resolve these issues.

Establishing private space

One way to evaluate privacy is to consider both technological and psychological privacy (Frankel & Siang 1999). Data that is protected by a password, or only available to certain people in a social network, is technologically private. A psychological perspective appreciates that the providers of accessible information still might not view it as public. A combined approach would ‘develop a technological understanding of the issue and then [expand] this understanding to include the psychological perspective of the participants’ (Frankel & Siang 1999: 11). In order to evaluate the privacy of blogs postings it is therefore necessary to examine the online location in which they are situated. The host sites enable users to restrict postings to certain users and have published policies regarding the usage of data. Ford (2011) highlights the use of these features in the ‘active management of privacy’ amongst bloggers and users of social networking sites. In addition, some bloggers might make an active effort to hide their ‘real’ identity in their blogs so they are not recognisable to family and friends (Serfarty 2004, cited in Ford 2011: 558). According to the ethical guidelines of the AoIR, this suggests that blogs can be considered as a more public webpage than, for example, conversations in a closed chatroom (Ess et al. 2002: 5; 7). So, in terms of technological privacy, the gap year blogs were public. However, if this is combined with perceptions of privacy, then this might be different. Would a gap year blogger expect anyone they did not know to be reading, not to mention analysing, their blogs?

There is considerable empirical research which suggests that there can be a lack of awareness about who can access online communications, and that people may act as if public online environments are private (Hudson & Bruckman 2005: 298). It may seem obvious that blogs that are not protected can be viewed by anyone, but this mediated environment can lead to ‘feelings of anonymity online or the ephemerality of text in chat or the (in)visibility of audience in blogs’ (Hudson & Bruckman 2005: 299). In order to address this, Wilkinson and Thelwall (2011) consider the difference between natural and normative privacy as put forward by Moor (2004). Natural privacy refers to situations when we would expect to be hidden, such as in a remote place, where any intrusion would be an accidental occurrence. Normative privacy refers to situations when we would expect others to protect us; it is reasonable to expect that a retailer will protect our banking information (Wilkinson & Thelwall 2011: 396). From this perspective, Wilkinson and Thelwall suggest that there is no reasonable expectation that personal information should be protected by others in a public web page, and suggest that the issue is one of natural, rather than normative privacy. Accessing personal information on the public web is:

...not a violation of privacy, but more of an accidental occurrence and one that the individual could not expect to have others protect them from. Hence, it seems reasonable, in principle, to use personal data from the public web for research purposes (Wilkinson & Thelwall 2011: 396).

Some researchers have suggested that it is not the responsibility of researchers to protect the information put online. Walther (2002) argues that if participants have misplaced expectations of privacy, the public needs to be educated about the accessibility of their online communications. For Walther, this is more important than whether researchers should be able to use them (2002, cited in Orton-Johnson 2010: 4.2). However, Ford (2011) suggests nothing is wholly public or wholly private, but needs to be considered on a continuum, with ‘an infinite and an infinitely variable number of configurations’ (Ford 2011: 561). Moreover, issues of privacy do not only apply to the information itself, but what it is used for (Lomborg 2013: 23). Both Lomborg (2013) and Wilkinson and Thelwall (2011) point to Nissebaum’s (2010) work in this area, which stresses the importance of context in terms of information being used in an appropriate way. The principle of contextual integrity (Nissebaum 2010) is considered by Wilkinson and Thelwall as a way of critically engaging with the simple distinction between natural and normative privacy. This takes into account the reactions of users to the use of their personal information in ways that are seen to be inappropriate. However, they suggest it seems unlikely that unobtrusive academic research would have a direct impact upon users that would prompt any negative reactions, as long as it is reported appropriately (Wilkinson & Thelwall 2011: 396-7).

Thus far, the gap year blogs could be considered as neither technically nor normatively private, but it seemed important to assess the bloggers’ perceptions of privacy and my responsibility towards them in this particular research context. Deciding about this responsibility rested in part on whether the gap year bloggers were classed as research subjects, whom needed protection, or authors, who had published their work online.

Bloggers: subjects or authors?

As noted above, one of the guiding principles of research ethics is ensuring that no harm comes to human subjects. How to establish whether subjects are involved is a common topic for online research ethics, as the internet ‘complicates the fundamental research question of personhood. Is an avatar a person? Is one’s digital information an extension of the self?’ (Markham & Buchanan 2012: 7). The question for the gap year study was whether these personal blogs are representations of human subjects, or are texts produced by authors (Lomborg 2013: 21-22). On one hand, it can be argued that ‘those that we encounter online should be categorised as human participants’ (Grinyer 2007: 2). Rather than a ephemeral ‘cyberspace’ where self-

presentation is detached from ‘bodies, faces or histories’, online communications reflect an offline ‘real life’ (Grinyer 2007: 2). Treating bloggers as participants would necessitate adhering to ethical guidelines regarding the prevention of harm to subjects of research. In the gap year study, determining whether the bloggers were subjects or authors had implications for two crucial decisions: whether informed consent was required to use the blogs as data, and how to report the data when writing up.

Informed consent

Firstly, informed consent is based upon the principle that individuals have the right to be briefed about what taking part in research will involve so that they can decide whether they wish to participate. Again, assessing privacy was a key factor. According to the original AoIR guidance ‘the greater the acknowledged publicity of the venue, the less obligation there may be to protect individual privacy, confidentiality, right to informed consent, etc’ (Ess et al. 2002: 5). Human subject research norms such as informed consent do not apply to public, published material. Consent is not required to analyse traditional texts such as newspapers or literature. Furthermore, sending participation requests may in itself be seen as intrusive. Hewson et al (2003) question whether contacting potential participants may be viewed as ‘spamming’, an invasion of privacy (Hewson et al 2003: 40). This raises the question of whether the interaction with the bloggers as part of the gap year study was appropriate.

Analysis and dissemination

There are also ethical implications regarding the analysis and dissemination of unsolicited online material. Published documents can be analysed even if there is a negative impact on the author or they do not agree with the conclusions, for example in the case of literary criticism (Wilkinson & Thelwall 2011: 394). In research involving human subjects however, harm may come to participants if their identity is disclosed, and researchers are required to employ measures to prevent this. The difficulty with making internet data anonymous is its ‘traceability’: ‘the possibility of locating digital data on the internet using search engines or any other mechanisms enabled by digital platforms (log files, user profiles, etc)’ (Beaulieu & Estalella 2012: 34). If a quotation taken from a blog post is put into a search engine, the origin of the text can be easily found and the creators identified. Linking data to its origin in this way is a potential breach of privacy, and thus it could be argued that it is necessary to avoid including identifying information (Wilkinson & Thelwall 2011: 397).

Lomborg (2013) cites the ‘distance principle’ in order to assess the applicability of human subject research norms. This is the relationship between an individual’s activities and the resulting data. For example, there is a considerable distance between a data set comprised of millions of log files and the people whose interactions created this data. There is a much smaller distance between a social networking profile page and the person who created it. Methods of data analysis are also important here: presenting data in aggregate puts greater distance between individuals and the resulting findings than a close content analysis of a text. Lomborg (2013) suggests that:

... we may more likely identify a study using a qualitative sample of ... blog posts as one that involves human subjects, because the experiential distance between an identifiable blogger and a blog text is perceived to be small (2013: 22-23).

From Lomborg’s perspective, the gap year bloggers *were* human subjects. However, to apply anonymity without establishing consent would take ownership of the blog data away from the bloggers for use by the researcher. This raises political questions. As Moore notes, ‘anonymisation allows the researcher to abstract the data from the research participants – and attach it to their own name in the process of publication’ (2012:

336). Some scholars have also contested the application of the human subject model to online communications.

How useful is the human subject model?

White (2002) argues that conceiving internet material as ‘people’ rather than ‘representations’ does not engage with insights from the Humanities, and points to how cultural disciplines could offer critical interventions in conceptualising this material as mediated and constructed, rather than ‘natural’. Alternatively, if the blog producers are treated as authors of documents then issues of ownership of material must be considered. Bassett and O’Riordan (2002) contest whether it is appropriate to apply the human subject model to internet communications. Instead, we can see the internet as a site for the ‘cultural production of texts’ (Bassett & O’Riordan 2002: 235), and the producers of online content as authors. If authors have chosen to publish in public, their authorship should be recognised. This would involve the citation of texts, just like traditional media (Bassett & O’Riordan 2002: 244). Taking this perspective would mean that every time a blog was referred to, a link would be provided to the post in question. If blogs are viewed as documents then the issue is one of copyright rather than breaching privacy (Wilkinson & Thelwall 2011: 395). In considering the methodological and ethical issues raised by blog research, Hookway (2008) identifies this as a key tension: whether researchers need to preserve anonymity or acknowledge authorship (Hookway 2008: 106). A ‘feminist ethics-of-care’ rather than paternalistic protection may be one way to negotiate such tensions on a case by case basis (Moore 2012: 338).

The recently updated AoIR guidelines suggest that assessing the principles related to human subjects may not be that helpful for research that does not interact with people directly, as people are always involved in the production of digital information at some point (Markham & Buchanan 2012: 6). Assessing personhood in online communications ‘may oversimplify the situation – the question of whether one is dealing with a human subject is different from the question about whether information is linked to individuals’ (Markham & Buchanan 2012: 7). I will return to this final point in the concluding section to critically reflect on the decisions made in the gap year study. Before I do so, I will now outline the decisions made in light of the debates presented.

Ethical decision-making in context

Two tensions: public versus private, and subject versus author, were central when making ethical decisions in the gap year study. Resolving these meant a very careful consideration of context in order to ‘think through’ how to apply the ethical guidelines to the data in question. As discussed above, advice from internet scholars emphasises the importance of context in making ethical decisions. Issues to take into account included expectations of privacy and the status of the bloggers in relation to their blogs, along with the methods of data collection and analysis, and how the data was to be used.

Firstly, the division between public and private in online spaces has to be considered as contextually dependent. Some useful advice from Whiteman (2007) suggests reading online communications to assess whether there is any awareness of unknown readers. This seemed to be the case for at least some of the gap year blogs. There were indications in the interviews that certain things were not discussed. One young woman commented that she was aware that the vicar was reading her blog. Another interviewee interacted with his closest friends via Facebook, and kept his blog more formal. There were also signs in the blogs themselves that the audience may be people who were unconnected to the blogger, so they provided a short biography or introductory post for ‘random readers’. In these examples, there is a potential case for considering blog data in the public domain, as it is not only publicly accessible but written for an audience.

The second key factor was to assess if the research involved human subjects. The ESRC Framework for Research Ethics defines human participants as living humans (which the bloggers certainly were) but that material already in public domain is not human data. This includes ‘information provided in forums or spaces on the internet and web that are intentionally public’, although this should be critically examined (ESRC 2012: 11). Internet scholars also point to the US Government’s Policy for Protection of Human Research Subjects:

‘a living individual about whom an investigator... conducting research obtains (1) Data through intervention or interaction with the individual, or (2) Identifiable private information’ (cited in Frankel & Siang 1999: 16).

The blog data were not produced through interaction and were not dealing with private information. Thus, the project did not fall under traditional guidelines for human subjects research. It is possible to make an argument that it is reasonable to use this without consent.

Given the bloggers were viewed as authors, and it was their public documents that were the object of study, issues of ownership then needed to be dealt with. The gap year study took a qualitative approach, with each blog as a unit of analysis. In other words, the data was not presented in aggregate, but was a thematic analysis of the texts, which would traditionally provide quotations from the source. Viewing the bloggers as published authors would mean that quotations would be referenced through citation to the source blogs. This is not necessarily a suitable approach, as the distinction between ‘subject’ and ‘author’ is not clear-cut. Young (2013) argues that even unobtrusive internet research should consider online texts as extensions of persons, particularly if the focus is on the content and tone, rather than, say, its linguistic structure.

As semi-structured interviews were also conducted with nine of the bloggers, ethical guidelines were followed to obtain informed consent for this phase of the research. Consent was also obtained to quote from interviewees’ blogs. This is because in the interview context, the interviewees were obviously ‘human subjects’ as data was generated by research-participant interaction. Interestingly the interviewees did not show any concerns or issues with the use of their blogs in an academic study; rather a mild curiosity about how I had found their blogs. The fact that the participants were not perturbed by the ethical approach taken suggests that this was an appropriate course of action for this stage of the research.

It was also important to remember that younger people may be more open online, and put surprisingly detailed personal information on their blogs (Huffaker 2006). There were certain personal details provided on the gappers’ blogs, and it was necessary to make a judgement regarding whose responsibility it was to protect this. While the bloggers had made this ‘public’, citing the blogs in any dissemination would mean that their identity would be easy to for any readers to uncover. Taking this into consideration, weblinks were not provided when writing up the research. This attempts to remove the *direct* links between the data and the person who has produced it. Instead of one policy on disguising individuals in online research, Bruckman (2002) suggests a ‘continuum of possibilities’ (Bruckman 2002: 229). This continuum includes various levels of disguise when reporting research, and the gap year study adopts ‘moderate disguise’ (Bruckman 2002). Personal details, such as names, blogger pseudonyms, identifiable details and specific locations were changed, although verbatim quotations were used. The traceability of digital information means that the blogs could be found by entering these quotations into a search engine, but they could also be found by searching for blogs on gap years. When resolving the tension he describes between preserving anonymity and acknowledging authorship in his own study, Hookway (2008) also takes this approach, and privileges the protection of his participants’ identity (2008: 106).

Evaluating the gap year study

I have put forward the case for a context-based approach to internet research ethics. Of course, this does not mean that 'anything goes'. Instead, those interested in online research ethics need to critically reflect upon and evaluate the decisions they make, and draw on the scholarship of others. At the time of the study, the decisions described above seemed reasonable and apposite. I do not think there will have been any harmful outcomes for the gap year bloggers whose blogs were included in my sample. In the process of continuing to disseminate my findings to a wider audience, and in following the continuing debates in internet research, I have reflected on whether the ethical tensions in the study were negotiated in the most appropriate way.

Given that notions of the public / private distinction have shifted, the concept of the 'public domain' is not necessarily helpful in making these ethical decisions. The blogs may address random readers, but not 'random researchers'. Using them in this way could encroach on the 'intimate publics' (Berlant 2008) that surround blogs. This highlights Reid's (1996) distinction between the ethics of access and the ethics of use (cited in Lees 2000: 13). Pauwel asks: 'Is using these Websites in radically other ways than those intended by their creators unethical?' (2006: 366). Consequently, a case could be made that consent should have been obtained for using the blogs as data.

Young (2013) suggests it is best err on the side of caution and obtain consent. In the case of Lomborg's (2013) research that utilised the personal archives of individuals in Denmark, including blogs and social network presence, she felt 'ethically obliged to ask permission' from the creators of this content (2013: 25). Lomborg assessed this content as experientially close to their everyday life, plus their expectations of privacy could not be known in advance and anonymity would have been impossible due to the traceability of the data. Consequently, Lomborg wanted to share her analysis with her participants. This approach can be seen as part of the wider rationale for Lomborg's study, which was built upon a dialogue with the producers of the personal archives, and the cultural context of Danish cultural and social norms, feminist approaches to research, and seeing informed consent as a continuous negotiation (Lomborg 2013).

Alternatively, making contact with the bloggers is potentially problematic. Wilkinson and Thelwall (2011) suggest that obtaining informed consent for publically available documents may not be appropriate. They provide an example of studying suicide-related online discussions; would it be appropriate to contact individuals in these circumstances? This could not only be conceived as an intrusion of privacy, but indeed might be potentially harmful (Wilkinson & Thelwall 2011: 395). Such concerns raise another point regarding the gap year study: all of the gappers were contacted for interview. Interacting with the bloggers, even if it was just via an interview request that was ignored and not responded to, involved them in the research in some way; they became participants. An alternative option would have been to conduct the research completely unobtrusively, and to not contact the bloggers at all.

Another option would have been to provide links to the blogs. This would acknowledge the bloggers' status as authors. While copyright legislation allows for reproducing works for the purposes of non-commercial research under 'fair dealing', this usually requires 'sufficient acknowledgement' (UK Intellectual Property Office 2013). There may also be a moral and political imperative to provide credit. In discussing the political issue of attribution in feminist blogging, Hvizdak (2008) notes: 'creation in the blogosphere depends heavily on the linking and quoting of others' information and the subsequent trust that proper attribution will be given' (2008: 130). By not providing a link to a blog, but using it without consent, could go against the cultural norms of using content in the 'blogosphere'. The approach taken by Leggatt-Cook and Chamberlain (2012) in their study of weight-loss blogs was to cite the blogs they used. All of the blogs in their sample were public, and indeed aimed to attract readers, so consent was not obtained, and quotations were used verbatim. The researchers' ethical judgement was used to ensure that the bloggers were represented in a sensitive way. There

are parallels here with the politics of anonymity in archival research. Moore argues that default anonymisation means that questions of naming and alternative approaches are not considered. She suggests it may be useful to explore the practices of historians who consult documents that were not intended for archiving or reuse (Moore 2012: 336). This would be a potentially useful avenue for internet researchers to investigate when evaluating how to deal with anonymity.

A final option for the gap year study would be to paraphrase the blogs, rather than quote them directly. This is another option on Bruckman's (2002) continuum of possibilities. This approach is advocated by Wilkinson and Thelwall (2011) when conducting unobtrusive research on personal information. Disguising the information addresses the issue of traceability: 'to avoid including identifying information in research publications and to avoid quotes or anonymize them by paraphrasing or altering words so they are not searchable' (Wilkinson & Thelwall 2010: 397). Returning to the AoIR guidelines, they advocate that it be appropriate to consider the potential harm of linking data to individuals, rather than deciding if human subjects or texts are the data source (Markham & Buchanan 2012: 7).

Conclusion

Innovative methods such as blog analysis present a number of ethical challenges. This paper has outlined the case for a contextual approach to internet research ethics in order to meet these challenges, in line with contemporary work in the field. In keeping with established principles, I argue that traditional ethical considerations need to apply to such methods, but that the blurred boundaries presented by the online, mediated environment means they require careful application. The gap year project used to illustrate this discussion approached this initially as unobtrusive research to obtain narratives of experience. Two key concerns were addressed: whether the blogs were public or private, and whether the bloggers should be considered as subjects or authors. During the project, the gap year blogs were assessed as in the public domain so informed consent was not obtained, but relatively young people had produced them and provided certain personal information within them. Consequently, a case was made to use the principles of 'moderate disguise' for the bloggers' identities (Bruckman 2002), rather than recognising their status as authors.

In reflecting on these ethical decisions, I have reviewed alternative approaches that could be taken. On reflection, drawing on the updated AoIR guidelines (Markham & Buchanan 2012), I have considered instead how research data may link to individuals, and how this should influence decisions in analysis and dissemination. Looking at the issue from this perspective would recognise that, while the blogs were the units of analysis, they were close to the young people who produced them. This means that, in order to address the issue of traceability in digital environments, it may have been more appropriate to paraphrase and fully anonymise the quotations taken from the blogs when consent could not be obtained. Although the bloggers who were interviewed gave their consent for quotations from their blogs, they could also have been asked if would like to be cited as authors. There is usually more than one ethically-defensible position on any given situation, but this is a more robust approach than the initial decisions². In presenting this research, I have critically reflected upon my own research practice in order to add to continuing discussions of how to conduct internet research in ethically sensitive ways that contribute to the development of critical social science.

Acknowledgements

This work was supported by the Economic and Social Research Council [grant number PTA-031-2004-00276]. I would like to thank the anonymous referees for their insightful comments on an earlier draft.

Biography

Helene Snee is Research Associate to the Head of Social Sciences at the University of Manchester. Her research interests include social stratification and socio-cultural change, and social interaction and practices, with a particular focus on young people's consumption, taste and cultural capital. She also has an interest in online methodologies and ethics, and how these illuminate wider questions for the social sciences. She completed her PhD at the University of Manchester in 2010, which examined the practice of gap year travel with reference to ideas about place, taste and self-development, and utilised weblogs as primary data. She is co-convenor of the BSA Youth Study Group (<http://www.britsoc.co.uk/study-groups/youth.aspx>) and was co-investigator on a ESRC National Centre for Research Methods Network for Methodological Innovation entitled Digital Methods as Mainstream Methodology (<http://digitalmethodsni.com/>). She has forthcoming publications on gap years as cosmopolitan practice and as identity work.

References

- Arioso, L. (2010) 'Personal Documents on the Internet: What's New and What's Old', *Journal of Comparative Research in Anthropology and Sociology*, 1(2) 23-38.
- Bassett, E. H. and O'Riordan, K. (2002) 'Ethics of Internet Research: Contesting the human subjects research model', *Ethics and Information Technology*, 4, 233-247.
- Beaulieu, A. and Estalella, A. (2012) 'Rethinking research ethics for mediated settings', *Information, Communication and Society*, 15(1) 23-42.
- Beer, D. and Burrows, R. (2007) 'Sociology and, of and in Web 2.0: Some Initial Considerations', *Sociological Research Online*, 12(5) Retrieved from: <http://www.socresonline.org.uk/12/5/17.html>
- Berlant, L. (2008) *The Female Complaint: The Unfinished Business of Sentimentality in American Culture*. Durham: Duke University Press.
- Bruckman, A. (2002) 'Studying the Amateur Artist: A perspective on disguising data collected in human subjects research on the Internet', *Ethics and Information Technology*, 4 217-231.
- BSA (2002) *Statement of Ethical Practice for the British Sociological Association* Retrieved from: <http://www.britsoc.co.uk/equality/Statement+Ethical+Practice.htm>
- ESRC (2012) *Framework for Research Ethics*. Retrieved from: <http://www.esrc.ac.uk/about-esrc/information/research-ethics.aspx>
- Ess, C. and the AoIR ethics working committee (2002) *Ethical decision-making and Internet Research: Recommendations from the AoIR Ethics Working Committee*. Retrieved from: <http://aoir.org/reports/ethics.pdf>
- Eynon, R., Fry, J. and Schroeder, R. (2008) 'The Ethics of Internet Research', in N. Fielding, R. M. Lee & G. Blank (eds.) *The SAGE Handbook of Online Research Methods*. London: SAGE

- Ford, S.M. (2011) 'Reconceptualising the public/private distinction in the age of information technology', *Information, Communication and Society*, 14(4) 550-567.
- Frankel, M. and Siang, S. (1999) 'Ethical and Legal Aspects of Human Subjects Research on the Internet', *American Association for the Advancement of Science Workshop Report*. Retrieved from: <http://www.aaas.org/spp/dspp/srfl/projects/intres.main.htm>
- Germann Molz, J. (2012) *Travel Connections Tourism, Technology and Togetherness in a Mobile World*. Abingdon: Routledge.
- Goffman, E. (1974) *Frame Analysis: An Essay on the Organization of Experience*. Boston: Northeastern University Press
- Grinyer, A. (2007) 'The ethics of Internet usage in health and personal narratives research', *Social Research Update* 49. University of Surrey.
- Heath, D. Kock, E., and Montoya, M (1999) 'Nodes and queries: Linking locations in networked fields of inquiry' *American Behavioural Scientist*, 43, 450-63.
- Heath, S. (2007) 'Widening the gap: pre-university gap years and the "economy of experience"', *British Journal of Sociology of Education*, 28(1) 89-103.
- Hewson C., Yule, P., Laurent, D., and Vogel, C. (2003) *Internet Research Methods: A Practical Guide for the Social and Behavioural Sciences*. London: SAGE.
- Hine, C. (2011) 'Internet Research and Unobtrusive Methods', *Social Research Update* 61, University of Surrey.
- Hookway, N. (2008) "'Entering the blogosphere": some strategies for using blogs in social research' *Qualitative Research*, 8(1) 91-113.
- Hudson, J.M. and Bruckman, A. (2005) 'Using Empirical Data to Reason about Internet Research Ethics', in H. Gellersen et al. (eds) *ECSCW 2005: Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work*, 18-22 September 2005. Paris, France.
- Huffaker, D. (2006). 'Teen Blogs Exposed: The Private Lives of Teens Made Public.' Presented to the *American Association for the Advancement of Science*, February 16-19, 2006. St. Louis, M.O., USA.
- Hvizdak, E. (2008) 'Creating a Web of Attribution in the Feminist Blogosphere', *International Journal of Internet Research Ethics*, 1(1) 115-134.
- Kelle, U. (1997) 'Theory Building in Qualitative Research and Computer Programs for the Management of Textual Data', *Sociological Research Online* 2(2). Retrieved from: <http://www.socresonline.org.uk/2/2/1.html>
- King, A. (2011) 'Minding the gap? Young people's accounts of taking a Gap Year as a form of identity work in higher education', *Journal of Youth Studies*, 14(3) 341-357.
- Lees, R. M. (2000) *Unobtrusive Methods in Social Research*. Buckingham: Oxford University Press.

- Leggatt-Cook, C. and Chamberlain, K. (2012) 'Blogging for weight loss: personal accountability, writing selves, and the weight-loss blogosphere' *Sociology of Health & Illness*, 34(7) 963–977.
- Lomborg, S. (2013) 'Personal internet archives and ethics', *Research Ethics*, 9(1) 20-31.
- McNeill, L. (2003) 'Teaching an Old Genre New Tricks: The Diary on the Internet', *Biography*, 26(1) 24-47.
- Madge, C. (2006) *Online Research Ethics*. Retrieved from:
<http://www.geog.le.ac.uk/orm/ethics/ethcontents.htm>
- Markham, A.N. and Buchanan, E.A and the AoIR ethics working committee (2012) *Ethical decision-making and Internet Research: Recommendations from the AoIR Ethics Working Committee*. Retrieved from:
<http://aoir.org/reports/ethics2.pdf>
- Miller, D. (2011) *Tales from Facebook*. Cambridge: Polity Press.
- Moor, J.H. (2004) 'Towards a theory of privacy for the information age' in R.A. Spinello and H.T. Tvani (eds.) *Readings in CyberEthics* (2nd ed) Sudbury, MA: Jones and Bartlett.
- Moore, N. (2012) 'The politics and ethics of naming: questioning anonymisation in (archival) research', *International Journal of Social Research Methodology*, 15(4): 331-340.
- Morrison, A. (2011) "'Suffused by Feeling and Affect": The Intimate Public of Personal Mommy Blogging' *Biography*, 34(1): 37-55.
- Nissenbaum, H. (2010) *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford: Stanford University Press.
- Orton-Johnson, K. (2010) 'Ethics in Online Research: Evaluating the ESRC Framework for Research Ethics Categorisation of Risk', *Sociological Research Online*, 15(4) 13. Retrieved from:
<http://www.socresonline.org.uk/15/4/13.html>.
- Pauwels, L. (2006) 'Ethics of Online (Visual) Research', *Visual Anthropology*, 19(3-4) 365-369.
- Reid, E. (1996) 'Informed Consent in the Study of On-line Communities: A Reflection on the Effects of Computer-Mediated Research', *Information Society*, 12, 169-174.
- Serfaty, V. (2004) Online diaries: towards a structural approach, *Journal of American Studies*, 38, 457–71.
- Simpson, K. (2004) 'Doing Development: The Gap Year, Volunteer-Tourists and a Popular Practice of Development', *Journal of International Development*, 16, 681–692.
- Snee, H. (2012) 'Youth Research in Web 2.0: A Case Study in Blog Analysis' in S. Heath and C. Walker (eds.) *Innovations in Researching Youth*. Basingstoke: Palgrave.
- Snee, H. (2013) 'Framing the Other: Representations of difference in overseas gap year narratives', *British Journal of Sociology*, 64(1)142–162.

UK Intellectual Property Office (2013) *Copyright*. Retrieved from:
<http://www.ipo.gov.uk/types/copy.htm>

Walther, J. (2002) 'Research ethics in Internet-enabled research: human subjects issues and methodological myopia', *Ethics and Information Technology*, 4(3) 205-16.

White, M. 'Representations or people?', *Ethics and Information Technology*, 4(3) 249-266.

Whiteman, N. (2007) *The Establishment, Maintenance and Destabilisation of Fandom: A Study of Two Online Communities and an Exploration of Issues Pertaining to Internet Research* PhD Thesis, Institute of Education, University of London.

Wilkinson, D. and Thelwall, M. (2011) 'Researching Personal Information on the Public Web: Methods and Ethics', *Social Science Computer Review*, 29(4) 387-401.

Young, K. (2013) 'Researching Young People's Online Spaces', in K. te Riele and R. Brooks (eds) *Negotiating Ethical Challenges in Youth Research*. New York: Routledge.

¹ A 'bottom-up' approach in this context's refers to the AoIR's decision to develop guidelines 'out of the day-to-day practices of researchers in a wide range of disciplines, countries and contexts, in contrast to a more usual top-down approach that tries to provide a universal set of norms, principles, practices, and regulations' (Markham and Buchanan 2012: 2).

² I am grateful to the anonymous reviewer who suggested this.