Received 7 October 2014
Revised 5 March 2015
Accepted 10 March 2015

Ethical considerations of using information obtained from online file sharing sites

The case of the piratebay

Aimee van Wynsberghe
Center for Telematics and Information Technology, University of Twente, Enschede, The Netherlands, and
Jeroen van der Ham
System and Network Engineering Research Group, University of Amsterdam, Amsterdam, The Netherlands

Abstract

Purpose – The purpose of this paper is to develop a novel approach for the ethical analysis of data collected from an online file-sharing site known as The PirateBay. Since the creation of Napster back in the late 1990s for the sharing and distribution of MP3 files across the Internet, the entertainment industry has struggled to deal with the regulation of information sharing at large. Added to the ethical questions of censorship and distributive justice are questions related to the use of data collected from such file-sharing sites for research purposes.

Design/methodology/approach – The approach is based on previous work analysing the use of data from online social networking sites and involves value analysis of the collection of data throughout the data's various life cycles.

Findings – This paper highlights the difficulties faced when attempting to apply a deontological or utilitarian approach to cases like the one used here. With this in mind, the authors point to a virtue ethics approach as a way to address ethical issues related to data sharing in the face of ever-changing data gathering and sharing practices.

Practical implications – This work is intended to provide a concrete approach for ethical data sharing practices in the domain of Internet security research.

Originality/value – The approach presented in this paper is a novel approach combining the insights from: the embedded values concept, value-sensitive design and the approach of the embedded ethicist.

Keywords Computer ethics, Design practice, IT ethics, Value-based design

Paper type Case study

1. Introduction

Since the creation of Napster back in the late 1990s for the sharing and distribution of MP3 files across the Internet, the entertainment industry has struggled to deal with the regulation of information sharing at large. From an ethics perspective, the practice of file sharing over the Internet presents an interesting value conflict between the protection of intellectual property (Von Lohmann, 2004), and fairness, or distributive justice (DeVoss and Porter, 2006). On the one hand, the entertainment industry wishes to uphold their exclusive copyrights of the content, to maintain their business model and their distribution methods. On the other hand, users are demanding easy access to music,

Journal of Information, Communication and Ethics in Society
Vol. 13 No. 3/4, 2015
pp. 256-267

1477-996X
DOI 10.1108/JICES-10-2014-0044

television and movie files, and will resort to file sharing when it is not available at a fair price or at all. Added to these kinds of ethical questions – at the macro level – are those related to the use of data collected from such file-sharing sites for research purposes, academic or otherwise – questions at the micro level. This paper aims to explore how the technological innovations related to file-sharing practices over the Internet have introduced ethical issues in the domain of Internet security research related to the collection and use of data from such sites. To do so, the authors develop a novel approach for ethical data provenance and the ethical analysis of data collected from an online file-sharing site known as The PirateBay (TPB).

The technical researcher of this paper was approached by the ISOC-NL working group on Internet transparency to assist in proving that the ban on TPB was not working and further to show that residents of The Netherlands were still accessing it (Poort *et al.*, 2014). To accomplish this, the researchers created a tool for observing the peer-to-peer file-sharing behaviour of users on TPB. The researchers were able to collect a great deal of information pertaining to individual users. The researchers approached the ethicist of this paper to assist them in understanding if they were right in creating such a tool and if they are allowed to share any of the data collected as a result of their methods.

The aim of this paper is three-fold:

(1) to conduct a retrospective ethical analysis of the collection and use of data obtained through monitoring online file sharing;

(2) to explore the role and utility of *ad hoc* ethics advice; and

(3) to suggest a framework for ethical data provenance (i.e. documenting the life cycle of the data alongside the ethical values and tensions that may arise).

The first two goals use the example of this TPB research as a case study for analysis. The third is a result of TPB analysis.

We begin by providing details about the kind of ethical analysis we engaged in, and we continue on to explain the technical details of TPB case study while analysing it. Although it is possible to question the utility of an *ad hoc* ethical analysis of this kind, we believe the methods and the findings to be of significant use for future researchers interested in ethics, as it pertains to:

- the use of data from online social networking sites;
- computer security research;
- data sharing practices; and
- data provenance concerns in general.

We conclude with suggestions and guidelines for future data sharing practices in information and communication technology (ICT) research.

2. A framework for ethical evaluation

To begin our assessment of the ethical issues, as they relate to the collection and use of data, we will use the framework developed by van Wynsberghe *et al.* (2013) for the ethical analysis of the collection of data from online social networking sites. In so doing, we do not wish to argue that TPB website resembles an online social networking site like Facebook or Twitter where all individuals are cognizant of their choice to share

personally identifiable information. Rather, we aim to show that there are certain similarities in the type of ethical questions asked when data are extracted from an online site in which personally identifiable information is shared.

The framework entails an analysis of decision variables and choices of the researcher rather than a study of the ethical intentions (Chen *et al.*, 2009) or decision-making choices of file sharers (Shang *et al.*, 2008). The framework consists of four components:

(1) the context of use and the privacy concerns for this context;

(2) the type and method of data collection;

(3) the intended use of information and the amount of information collected; and

(4) analysis of values.

The ethical analysis to come in the following sections entails a description of the technical details of the case study according to the above components while at the same time engaging in value analysis.

To give a short word on value analysis, we take here the starting point of the embedded values approach (Nissenbaum, 2001; Friedman, 1996) that claims there are values embedded into a technology so that when the technology is used the value is made real. The value is then a consequence of using the technology. With this in mind it is suggested that technologies can and should be made to intentionally realize desirable values while minimizing undesirable consequences (van Wynsberghe and Robbins, 2013). This conclusion assumes a prospective approach to technology design and in this instance we are engaging in a retrospective ethical appraisal. As such we will focus on the values embedded in this work and whether or not the intended values of the researchers were in fact realized in the methods used.

The first step in this type of value analysis (van Wynsberghe and Robbins, 2013) is to make the implicit values intended by the engineers explicit. By making explicit the intended values it is possible to do a variety of things:

- scrutinize these values;
- uncover any value trade-offs;
- compare these values with the ethics literature; and
- question whether or not these values were in fact realized in the final outcome of the research case study.

The following sections will do just that and will conclude with overall recommendations for future ethical analyses of data sharing practices.

3. The project: the case of TPB

For the reader to follow the line of thought in the following sections, we would first like to say a few words on the overall research methodology and why it was initiated. The study is related to TPB, which is a website that facilitates the sharing of entertainment files like movies, songs, television programs, etc. Users can have access to this material for free when using links from such a website. This presents a problem for the entertainment industry whose revenues come from paying consumers/citizens. In different countries, the entertainment industry lobbying organizations are taking

different approaches to combat this issue. For a variety of European countries, access to TPB is blocked, but users are finding ways around this.

In The Netherlands, the entertainment industry successfully won a court case in 2009 against TPB forcing them to block users from The Netherlands. The website, however, has not recognized the ruling and has not taken action to block any users in The Netherlands from accessing the site. By 2012, BREIN sued Internet service providers (ISPs) to implement a blockade for TPB. At first, it was only for two ISPs (Ziggo and XS4ALL), but a few months later, most of the remaining ISPs were included. This situation provided a unique opportunity to study the effects of a website blockade on the file-sharing behaviour of consumers. To that end, van der Ham *et al.* (2012), Poort *et al.* (2014) began to measure whether preventing access to these sources of links had an impact on file-sharing behaviour of users in The Netherlands. The hypothesis was formed that if the blockade of the TPB website was effective, then there would be a significant difference in BitTorrent users of TPB files before and after a new blockade.

3.1 The context of the swarm and the meaning of privacy

The significance of the context from which information or data are collected has been stressed by computer ethicist Nissenbaum (2009) as a foundational component to the ethical analysis. The context from which data are collected or in which research is conducted is important for a variety of reasons:

- It is directly related to how the value of privacy is conceptualized and prioritized.
- It is directly related to the kind and amount of information that can be collected, to name a few (van Wynsberghe *et al.*, 2013; van Wynsberghe and Robbins, 2013).

For these reasons, it is important to discuss the context in which the research is taking place, or from which the data are being collected, in terms of the concept of privacy, as it relates to the data acquired.

The context of the research we are discussing here is TPB website which facilitates peer-to-peer file sharing. To give some background here, we will first describe the file-sharing mechanism: BitTorrent. BitTorrent is a peer-to-peer file-sharing protocol. Users of a BitTorrent client, "peers", collaborate in distributing a file to each other. To distribute a set of files, a "torrent" file is created, which provides some metadata necessary to be able to exchange the file. This torrent file or a special link, "magnet-link", to that file is published in some way, for example on the TPB website, where other users can download it. Using a BitTorrent client, the user loads the torrent file (directly or using the magnet link) and then finds peers who are also downloading those sets of files, together forming a "swarm". The peers cooperate in the swarm until they have the complete content and preferably longer to help others get the content. For more information, see Cohen (2003).

A user goes to TPB to discover content, and once the user clicks on the link, she/he joins the swarm. Note that the swarm does not include TPB itself. With this in mind, it is important to consider the concept of privacy, as it relates to the swarm rather than to the online site of TPB. Keep in mind, however, that it was TPB that allowed the user to access the swarm in the first place.

When discussing the value of privacy, traditionally, it refers to restricting access to certain personally identifiable information about a user. Such information might be a users' Internet protocol (IP) address or geographic location. What proves to be most

interesting for this context is that for the swarm to work (i.e. for something to be downloaded), the aggregate of computers making up the swarm *must* share information amongst each other. This information contains, but is not limited to, IP address (note that in The Netherlands IP, address is considered personally identifiable information). Thus, for the content to be shared accurately, the IP address must be shared. Consequently, in a context like this, there can be little to no expectation that information like IP address will be kept private from the other computers/users in the swarm.

That being said, one cannot conclude from this that such information can then be extracted and used for other purposes. That would qualify as an unauthorized secondary use of the information: a privacy violation (Zimmer, 2010; Smith *et al.*, 1996). Thus, the issue of privacy in this context (i.e. the swarm) is quite problematic in that personally identifiable information such as IP address must be shared amongst users, but this information can only be shared for the purposes of uploading and downloading content.

To complicate things further, the very act of sharing content is illegal in many countries; thus, having access to who is engaged in such activities is information that is very useful for many stakeholders who wish to bring charges against those who engage in such practices. This brings into light the relationship that this discussion shares with the discussion of copyright law versus distributive justice, as it relates to entertainment content. In other words, if you see the swarm as doing something wrong, are you allowed to use the information from the swarm to make it right?

3.2 Method and amount of data collected

From an ethics perspective, an understanding of the methods used for data collection and the amount of information collected are important components, as they relate not only to the concept of privacy but also to other values like fairness and/or transparency. When "extensive amounts of personally identifiable data are being collected and stored in databases" (Smith *et al.*, 1996, p. 172), this can constitute a privacy violation. These extensive amounts can refer to data maximization – when you take more than what you need – and are balanced with the concept of data accountability – when you take what you need and give up what you don't need. Additionally, when information is collected using methods that do not seek consent, one must consider threats to privacy but must also consider the value (or lack thereof) of transparency in the overall research approach/methods.

The tool the researchers created and used for monitoring, "Threepwood", was developed using an existing library, which implemented the BitTorrent protocol (Nordberg, 2011). The advantage of this library was that the monitor would not actively participate in sharing and downloading content. Thus, the researchers were not engaged in any illegal activity while collecting information. Furthermore, Threepwood could be implemented in a distributed fashion, using three different vantage points, while submitting results to a central server. This was done to rule out a possible bias that might result from looking at the swarm from one vantage point. This also made it feasible to run Threepwood over longer periods of time, without attracting unwarranted attention. Longer periods of time for monitoring helped to ensure accuracy. To ensure that the monitoring activity was aimed at Dutch users, Threepwood used the top ten video torrents with "NL" in their name.

The monitor joined each of the swarms and collected information on the peers. The type of data collected was limited to: the IP address, the timestamp and which swarm. To maintain accountability, the researchers took only the data that were required and gave up that which were not essential to their objective. The central server received the monitored data from the distributed monitors over encrypted connections. The data were then stored in a local database for further analysis. The local database was secured in a general manner, and only the two researchers involved had access to the data. Extracting the geolocation from the IP addresses was performed locally, but an outside service was used to perform the lookup from IP number to the registered ISP or other organization. Assistance from this outside service was necessary: there was no other way to acquire said information.

After running Threepwood for several days, the results were analysed. The researchers were surprised to learn that Threepwood was able to record a very large number of the peers connected to the different swarms, most likely even recording all users who were connected to those swarms over that time period. Although researchers were expecting to get a small sample of information, they found they had collected a considerable amount of data from a larger than expected sample size. By retrieving information from such a large sample size the researchers were now in a position to have an incredible amount of PII from a large number of users. It was this finding that prompted the researcher to seek the advice of an ethics adviser.

When analysing the methods used for data collection from an ethics perspective, one must address that the data were collected in a passive way; meaning, no consent was sought from the users whose information was being collected. Thus, the researchers were not being transparent to the users in their method of data collection. When asked about this issue, the researchers indicated that this approach was necessary to have accurate results and to be as objective as possible in the acquiring of results. We may speak of a value trade-off at this point: the values of objectiveness and effectiveness were chosen in favour of the value of transparency.

This detail further confuses the discussion on privacy because if one wishes to obtain consent from the users, their privacy would have to be breached to do so (i.e. having the IP address of a user does not necessarily mean you can contact them, for that you need to contact their ISP to obtain such information.).

Added to this is the fact that the researchers were not actually participating in the uploading or downloading of content. Information like IP addresses is only meant to be shared when a computer engages in the uploading and downloading protocol; however, in the NL (the context where the researchers reside), uploading is illegal and the researchers did not want to take part. The BitTorrent protocol includes mechanisms to promote sharing using a tit-for-tat mechanism, and there is ongoing research to reduce "free riding" (Zghaibeh and Harmantzis, 2008). With this in mind, the behaviour of the researchers proves to be ethically problematic insofar as one might suggest the researchers were dishonest in their collection methods. This raises an additional privacy violation, namely, improper access to personal information (Zimmer, 2010; Smith *et al.*, 1996). Thus, added to the risk for unauthorized secondary use mentioned above, we have the issue of how the information was accessed.

Consequently, the values of transparency and privacy are traded off for the values of objectiveness and accuracy; researchers did not want the users to influence the results

and therefore did not ask their permission to collect the data that they had improper access to.

3.3 Intended use of data collected

The intended use of data collected is significant, as it relates to the intended values and goals of the researchers. Addressing this aspect creates room for motivation and purpose of the research to factor into the ethical landscape (McBride, 2014). This component also encourages a discussion of the moral character of the researcher – a component we will argue in favour of later in this paper – which bears significance for the outcome of an ethical evaluation.

The data collected in this research project were done with the intention that it be used in a court case showing the effectiveness, or lack thereof, of a blockade on TPB. The data were also intended to be used as a platform for discussing the effectiveness, or lack thereof, of blocking websites in general. Accordingly, one intended use was for the legal domain and another was for an academic domain. For both of these uses, it was necessary that the data be accurate and objective. These values were thus embedded into Threepwood and made real through the research methods used by the researchers, as seen in the previous sections.

The next stage in the data's life cycle is storage: the collected data were kept in a secured environment during and after the review period so that reviewers or other researchers could verify the methodology. The reason for these measures is to ensure the privacy and protection of the sensitive data that had been collected. It was not lost on the researchers that they had attained highly sensitive data that required protection.

After this, the next stage in the data's life cycle is dissemination. What is important here is whether the data are made public and to what extent. In this case, the original data were not made public at all. Instead, the researchers analysed and aggregated the data they had collected into a table showing the distribution of the swarms. This table does not contain any PII, but only the number of users per ISP in each country. These distributions were then compared to previous distributions and also to a questionnaire on users' behaviours. This comparison was used to prove the hypothesis of the study (Poort *et al.*, 2014). The original data were kept for reproducibility of the experiment, but it was not released to protect the privacy and anonymity of the users.

4. Ethical analysis: duties, consequences or virtues?

With all of this information we can now analyse the life cycle of the data and address the ethical considerations on a broader level, namely, that of data provenance. To begin, we suggest that the research conducted was non-human subjects' research, as there is no interaction with the actual participants, and the information obtained may not be personally identifiable. The PII was not readily personally identifiable and had to go through a third party for identification.

Although PII was collected in a way that posed certain privacy violations, the information that was used in the final output of the researchers (i.e. the aggregate data) was done in a way that specifically and intentionally protected the privacy and anonymity of the individuals in the swarm. In other words, the researchers made every attempt to control the scope of the privacy violation (Roux and Falgoust, 2012).

This reveals a core ethical dilemma of this case study: the distinction between the lack of protection of privacy during data collection versus the protection of privacy during data dissemination/sharing. If privacy is protected in the final dissemination or sharing of the data, is the privacy violation during data collection defensible? Furthermore, if the ends for which the data are to be used are to promote a greater good, then can such ends justify the means of data collection (Roux and Falgoust, 2012)? In other words, are there conditions relevant to this case study that warrants the privacy violation?

So, how can we evaluate this case study? A deontological approach would say the researchers were wrong for not adhering to the duty to protect privacy at every stage of the research approach, whereas a consequentialist would claim that the consequences of the research proved beneficial to the users; therefore, the ends may justify the means.

In the constantly evolving environment of information systems, it seems impossible to (re)create and disseminate hard and fast rules that can be applied to the ever-changing technology or to rely on an analysis of the consequences that are impossible to adequately predict at the onset of a research project or data sharing practice. Furthermore, guidelines for data minimization (i.e. collecting only the minimum amount of data needed) and anonymization are not universally defined and constantly changing with new forms of gathering data.

With this in mind, one may rightly wonder whether the most appropriate avenue to pursue is neither a deontological nor a consequentialist approach but a virtue ethics approach that focuses on the development of the technical researcher as virtuous in his/her professional role (McBride, 2014; Vallor, 2010). For that reason, McBride (2012, p. 24) has proposed the ACTIVE approach to ethical analysis of technical systems which is "derived from virtue ethics and extends the goal of the PAPA model". The PAPA model addresses the ethical issues of privacy, accuracy, property and access. This approach has been criticised by some for narrowly restricting ethical analysis of information systems to issues of privacy (McBride, 2014). Instead what is needed is a tool to broaden the conceptualization of ethics beyond issues of privacy, as it relates to information systems. To that end, the individual researcher's intentions and actions need to be integrated into the ethical consideration of this case study.

We do not claim here that the ends of using the data can justify any kind of privacy violation encountered in the collection or dissemination of data. Such a claim would mean that anyone could justify deplorable means of collecting data if it is being used for "good", consider the collection of data by the NSA, Facebook or Google as examples. Furthermore, who determines what the "good" is and how it is achieved is debatable. Alternatively, we cannot claim that if data are collected with assurance to privacy but disseminated while violating privacy that it is ethical.

Instead, we want to suggest that the researchers acknowledge the threats to privacy through data collection but were clear they had no other alternative to collect accurate data. Most importantly, the researchers made every attempt to protect the privacy of swarm users in the dissemination of the data. They could have taken more data than they needed or published the original data rather than the aggregate data but they believed that this did not adhere to the professional standards of their role as researchers or to the intended goal of the research.

This example also invokes a discussion of the plausibility and utility of the concept of informed consent when dealing with online contexts and data that are not easily identifiable. When data are collected in the manner described here (i.e. with the

researchers being explicit about their intentions for the use, storage and amount of data collected), perhaps one may go so far as to suggest that informed consent may not be necessary. Such a suggestion is conditional on the fact that the data collected would never be disseminated or shared in its original form (only as anonymized aggregate data) and would presume that the researchers were explicit at the outset of their project and had a written affidavit that expressed these intentions. One may go even further to suggest that because these subjects were breaking the law, their consent is not necessary. We do not wish to explore this option as the laws were constantly changing so from one moment to the next, the same person could be considered a law breaker and then a law abider.

5. The value of ad-hoc ethics for future work

Collaborating together, the ethicist and computer scientist found numerous points where ethical considerations played a dominant role in the decisions made by the computer scientist without their explicit knowledge. This speaks to the very manner in which engineers work: they often consider the ethics of their choices without knowing what they are doing or being capable of criticizing their own perspectives. Added to this, by engaging in the retrospective assessment and evaluation of what the researchers had done and why, the ethicist and computer scientist were able to gain a better understanding of each others' thinking and arrive at suggestions for other researchers engaged in this type of work.

As a proposal for best practice in data sharing, the authors suggest documenting the data's provenance. This is done through identifying the various stages of the research approach and engaging in a value analysis at each of the research stages. This practice is a first step for ethical data provenance, documenting the ethical values associated with each of the data's life cycles. In general, these stages may be labelled as follows:

- overall concept and design of research proposed;
- collection of data;
- storage of data;
- analysis of data;
- verification of data;
- dissemination of result;
- data sharing; and
- non-storage or destruction of data.

The list of values to be used in the value analysis is drawn from a variety of sources both from the computer scientists (e.g. the proposals and objectives of their research) involved and the ethical literature relating to the research (Table I).

By tracking the data collection through various life cycles, it becomes possible to isolate the ethical issues related to one stage and/or another. Additionally, it allows for the researcher to reflect on their professional role and the motivation(s) driving their actions throughout each of the life cycles of the data. Working together, the ethicist engages in an interview style process with the computer scientist(s) at each of the stages to explore:

- what values are intended, how they are conceptualized and translated into technical variables; and
- what value trade-offs reveal themselves.

With respect to the presence of value trade-offs, the ethicist can help explore alternative approaches to fulfil the objectives of researchers while minimizing disparities between values. This may not always be possible, but in many instances even making such trade-offs explicit and transparent is a first step in making a change.

One of the questions of the researchers had to do with what they could and could not do with the data after it had been collected: the researchers feared the original data would be requested by the entertainment companies to bring charges against users in the swarm or net neutrality advocates to argue against future blockades. We suggest that by paying meticulous attention to and documenting the manner in which data are collected, stored and analysed, it is easier to arrive at the decision of whether it is permissible to share said data. In this instance, we would never advocate in favour of sharing the original data but would claim that the aggregate data are permissible only when the researchers can reasonably ensure anonymization. This point reiterates the significance of developing and evaluating the virtues of the researcher in his/her professional role.

6. Conclusion

The intended goal of the researchers was to test the effectiveness of the blockade on TPB website. To achieve this goal, there are corresponding values the researchers had intended. Most notably the researchers, just like any other researcher in the same field, wanted their work to be accurate, objective and fair. At the same time, they wanted to protect their own reputation as researchers and wanted to promote the privacy and anonymity of the users in the final dissemination of the results. The research methods used were designed in a way to achieve this:

- They collected information from users without their knowledge so as to ensure that users could not lie.
- They collected information in a way that prevented them from breaking the law.
- They limited the amount of information collected to only the data that were needed for the objectives of the project.
- They ensured the privacy and anonymity of users in the swarm when the data were disseminated.

When we consider the intended values of the researchers and how these values were translated into the technical details of the tool used for data collection, we can see value

Data life cycle	Overall concept and design of research proposed, collection of data, storage of data, analysis of data, verification of data, dissemination of result, data sharing and non-storage or destruction of data
Stakeholder	Academic researcher, industry researcher, lawyer, journalist, user, . . .
Intended values	Privacy, accuracy, reliability, efficiency, reputation, justice, beneficence, anonymity, . . .
Value analysis	Value tensions, value trade-offs, value conflicts, . . .

Table I. Framework for ethical data provenance

trade-offs. As mentioned, the value of transparency and privacy was oftentimes prioritized lower compared to the values of objectiveness, effectiveness and accuracy during data collection. Most interesting was the observation that although privacy of users in the swarm was violated through the data collection, the manner in which it was stored and disseminated was done to intentionally protect privacy and anonymity of the swarm.

Tracking values in this way – as a first step towards ethical data provenance – allowed us to draw out the core ethical dilemma of this case, namely, if privacy is protected in the dissemination of the data, can this justify violations to privacy in the collection of data? We report, for this specific instance, that given the steps taken to ensure the confidentiality of the data (e.g. data storage and dissemination) and to control the scope of the privacy violation (e.g. publishing aggregate data rather than original data), the method of data collection versus the final protection of privacy is warranted.

Moreover, tracking values in this way also allowed us to delineate the responsibilities of the researchers for their actions with respect to data collection and dissemination. What's more, the additional proposed use of virtue ethics in evaluating ICT practices provides a broader perspective, namely, evaluating the motivations and actions of researchers. Consequently, the work provided, and argued for, in this paper offers a clear and all encompassing approach for data provenance with ethics at its core.

References

Chen, Y., Shang, R. and Lin, A. (2009), "The intention to download music files in a P2P environment: consumption value, fashion, and ethical decision perspectives", *Electronic Commerce Research and Applications*, Vol. 7 No. 4, pp. 411-422.

Cohen, B. (2003), "Incentives build robustness in BitTorrent", 1st Workshop on Economics of Peer-to-Peer Systems, Berkeley, CA, 5-6 June, pp. 68-72.

DeVoss, D.N. and Porter, J.E. (2006), "Why Napster matters to writing: filesharing as a new ethic of digital delivery", *Computers and Composition*, Vol. 23 No. 2, pp. 178-210.

Friedman, B. (1996), "Value-sensitive design", *Interactions*, Vol. 3 No. 6, pp. 16-23.

McBride, N. (2014), "ACTIVE ethics: an information systems ethics for the internet age", *Journal of Information, Communication and Ethics in Society*, Vol. 12 No. 1, pp. 21-44.

Nissenbaum, H. (2001), "How computer systems embody values", *Computer*, Vol. 34 No. 3, pp. 120-119.

Nissenbaum, H. (2009), "Privacy in Context: Technology, Policy, and the Integrity of Social Life", Stanford University Press, Redwood city, CA.

Nordberg, A. (2011), "Rasterbar libtorrent", available at: www.rasterbar.com/products/libttorrent

Poort, J., Leenheer, J., van der Ham, J. and Dumitru, C. (2014), "Baywatch: two approaches to measure the effects of blocking access to The Pirate Bay", *Telecommunications Policy*, Vol. 38 No. 4, pp. 383-392.

Roux, B. and Falgoust, M. (2012), "Ethical issues raised by data acquisition methods in digital forensics research", *Journal of Information Ethics*, Vol. 21 No. 1, pp. 40-60.

Shang, R., Chen, Y. and Chen, P. (2008), "Ethical decisions about sharing music files in the P2P environment", *Journal of Business Ethics*, Vol. 80 No. 2, pp. 349-365.

Smith, H.J., Milberg, S.J. and Burke, S.J. (1996), "Information privacy: measuring individuals' concerns about organizational practices", *MIS Quarterly*, Vol. 20 No. 2, pp. 167-196.

Vallor, S. (2010), "Social networking technology and the virtues", *Ethics and Information Technology*, Vol. 12 No. 2, pp. 157-170.

van der Ham, J., Rood, H., Dumitru, C., Koning, R., Sijm, N. and De Laat, C. (2012), "Review en herhaling BREIN steekproeven 7-9 april 2012", SNE Technical Report SNE-UVA-2012-01, Amsterdam.

van Wynsberghe, A., Been, H. and Keulen, M. (2013), "To use or not to use: guidelines for researchers using data from online social networking sites".

van Wynsberghe, A. and Robbins, S. (2013), "Ethicist as designer: a pragmatic approach to ethics in the lab", *Science and Engineering Ethics*, Vol. 20 No. 4, pp. 947-961.

Von Lohmann, F. (2004), "Measuring the digital millennium against the Darknet: implications for the regulation of technological protection measures", *Loyola Los Angeles Entertainment Law Review*, Vol. 24 No. 4, p. 635.

Zghaibeh, M. and Harmantzis, F.C. (2008), "Revisiting free riding and the Tit-for-Tat in BitTorrent: a measurement study", *Peer-to-Peer Networking and Applications*, Vol. 1 No. 2, pp. 162-173.

Zimmer, M. (2010), "But the data is already public: on the ethics of research in Facebook", *Ethics and Information Technology*, Vol. 12 No. 4, pp. 313-325.

Further reading

Larsson, S., Svensson, M., de Kaminski, M., Rönkkö, K. and Olsson, J.A. (2012), "Law, norms, piracy and online anonymity: practices of de-identification in the global file sharing community", *Journal of Research in Interactive Marketing*, Vol. 6 No. 4, pp. 260-280.

Mason, R.O. (1986), "Four ethical issues of the information age", *MIS Quaterly*, Vol. 10 No. 1, pp. 5-12, available at: www.ida.liu.se~TIMM32/docs/4etical.pdf

Corresponding author

Aimee van Wynsberghe can be contacted at: aimeevanrobot@gmail.com