To use or not to use: guidelines for researchers using data from online social networking sites.

Aimee van Wynsberghe[1], Henry Been[2], Maurice van Keulen[3]

Abstract

The universal and ubiquitous use of computing technologies confronts us with new ethical dilemmas on a daily basis. In the current age of abundant information sharing and gathering, social networking sites (SNSs) are now thought of as incredible resources for collecting data on individuals. The question of interest for computer scientists, engineers and ethicists alike asks: what are the ethical limits for using data obtained from online SNSs? To date, no practical guidelines exist for assisting researchers in either experiments or in the creation of products using data obtained from online SNSs. The following paper gives a general discussion of the main ethical considerations and proposes guidelines for incorporating ethics into research involving data from online SNSs. To demonstrate the practical relevance of the proposed guidelines a case study looking at online resolution identity using Twitter is used. This case study is of particular interest, because the research and the intended technology pose different threats. The suggestions presented here are also meant to stimulate the discussion and debate on the current and future use of such data and the creation of best practices.

Introduction

In the current age of abundant information sharing and gathering, social networking sites (SNSs) are now thought of as incredible resources for collecting data on individuals. To date, such data is collected in a variety of ways (e.g. passively or aggressively), by a variety of researchers (e.g. academic, industry, governmental) for a variety of purposes (e.g. detecting fraudulent behaviors, detecting consumer patterns, studying user patterns). Given this range in collection methods and uses of the data, the question of importance for ethicists, researchers and citizens alike has to do

[1] Ethics Advisor for CTIT (Center for Telematics and Information Technology), University of Twente, the Netherlands. Post Doctoral Researcher, University of Twente, the Netherlands.
[2] Student of Electrical Engineering, Mathematics and Computer Science (EEMCS), University of Twente.
[3] Associate Professor Data Management Technology. Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), University of Twente.

with when and for what can such data be used? In other words, what are the ethical limits when using data obtained from online SNSs?

Current philosophers of technology and computer ethicists are investigating the notion of embedded values – a technology is created in a non-neutral way such that through its use it may promote or threaten the realization of certain values (Nissenbaum, 2001; Friedman et al., 2006; Brey, 2000; Latour, 1992). From this foundation have come a plethora of design studies attempting to incorporate values into the design process of technologies (Nissenbaum, 2011; Friedman et al., 2006; Brey, 2000; van Wynsberghe 2012, 2012a). The greatest challenge facing value analyses to date is how to understand and conceptualize values in context (Manders-Huits, Noëmi, 2011; Van de Poel and Kroes, forthcoming), especially new online contexts like Facebook and Twitter (Nissenbaum, 2011), to understand the ethical limits of such research endeavors (Zimmer, 2012). Note that not only the intended technology may pose a threat, but also experiments conducted in the design or validation of the technology. Added to this is how to make the values and intentions of computer engineers and scientists explicit in order to: critically evaluate such values, uncover value trade-offs and track the realization of values when a technology is used or an experiment conducted. The goal in taking these ethical issues into consideration is two-fold: 1. to track value trajectories over time and, 2. to encourage the design of better products.

Given the lack of best practices in terms of applying ethics to the use of data from SNSs, this paper will present guidelines for researchers wishing to conduct, or already conducting, such research. These guidelines are created using suggestions from the current literature as well as first-hand experience as an ethics advisor for a research institute dealing specifically with the research and design of ICT systems. A case study looking at online resolution identity using Twitter is used to give a practical example of how to use such guidelines. This case study is of particular interest, because the research and the intended technology pose different threats. The suggestions presented here are also meant to stimulate the discussion and debate on the use of such data and the creation of best practices.

Main Ethical Issues Faced By Research That Uses Data from SNSs

Results from the Facebook "Tastes, Ties and Time" (T3) project initiated in 2006 revealed numerous important findings related to the ethical issues facing the collection and use of data from online SNSs for research purposes. In general, although researchers attempted to protect the anonymity and privacy of individuals, this was not achieved in practice. Using the dimensions for privacy violations of Smith et al. (1996), Zimmer was able to uncover four specific privacy violations (2012). For starters, too much information was collected. Second, information was collected in an unethical manner (passive collection of data) especially given the sensitive nature of the data. Third, information was used for purposes that had not been indicated in the original collection of data. And fourth, subjects were never given the opportunity to correct errors in the information that was gathered. The above discussion shows that the value of privacy is an issue

of extreme ethical importance that requires significant foresight and sensitivity in planning research initiatives using data from online SNSs.

Privacy is dependent on the context from which the information will be collected as well as the kind of information being collected. As Nissenbaum (2004) describes there are types of information that are intrinsically more sensitive in nature compared with other types. Information of this kind refers to one's name, sexual orientation or preference, age, and so on. This type of information makes it easier to identify an individual or can be quite revealing about an individual. It is therefore deserving of greater privacy consideration than other information like what an individual ate for lunch on a given day for example. There is also information that is more sensitive in nature given the context of where the information is generated or may be obtained. Information in the homes of individuals fits this profile: citizens are protected in their homes from interference by the government or other citizens. There are of course limits to how far privacy extends when talking about the kind of information above but the restrictions on obtaining, or using, information of a sensitive nature or from a certain context are lifted only when an individual is a suspect in a crime, e.g. fraudulent behavior.

Thus, it is crucial to understand the context from which the online data is obtained and the type of data (whether it is sensitive in nature). With respect to the first point, how can we discuss context in terms of online SNSs? One distinction is the intention of the owner/poster of the data. For example, generally, information on Facebook is meant to be shared with Facebook "friends" – individuals who have been granted permission to access it. On the other hand, Twitter is all about sharing with the world. In fact, in 2010, the Library of Congress announced that, "Every public tweet, ever, since Twitter's inception in March 2006, will be archived digitally at the Library of Congress" (Raymond, 2010). This kind of initiative would never be allowed with Facebook given that the intention of users is quite different. Added to this, it is also important to consider vulnerable groups (e.g. cyber bully victims) and/or those who do not understand how much information they are sharing about themselves on SNSs.

Referring back to the Facebook T3 project and the indicated privacy violations, the method in which the data was collected was argued to constitute a privacy violation (Zimmer, 2012). Data was passively collected meaning that subjects were not asked to participate; rather, their data was collected/scraped from their profile. When data is collected in this passive manner there is no option for informed consent or for correcting any errors in the data. When data is actively collected this refers to a participant's voluntary participation through informed consent forms or other types of agreement. There are obstacles to this practice as well given that it is not possible in all cases to identify exactly what the data will be used for when additional analyses may be completed that were not originally intended but that are required for proper results.

This last point raises the issue of how much information is collected and what this information will be used for. Even when it is not entirely possible to indicate every single use for the data, it could, and should, still become common practice to stipulate the limits within which a researcher

may work. For example, the researcher may say the information from Twitter will be used to test a model of identity resolution which may be used for the detection of fraudulent behavior but may not be used to ascertain commercial habits or proclivities of an individual for advertising purposes. With this, there are limits within which the researcher can work and the researcher and subject are well aware that the information will not be used for the specific purpose of marketing or advertising or for use 20 years from the data of collection, for example.

This discussion of ethical issues related to the collection and use of data from online SNSs is not exhaustive to say the least; however, it is intended to raise the most significant issues at this time so they may be addressed. The guidelines presented are intended to assist researchers in addressing these issues as well as pointing to additional ethical concerns that may arise in the design process.

Ethics and ICT in Practice

The methods used here are based on the work of an ethics adviser for a technical research institute, van Wynsberghe. Her work as ethics advisor for a CTIT (Centre for Telematics and Information Technology) at the university of Twente, is dedicated to the development of a pragmatic ethics to facilitate the incorporation of ethics into ICT research and design on a broad scale. An extensive outline of her methods and experiences are presented in another paper (see van Wynsberghe, 2013).

The tasks to be fulfilled by an ethics advisor differ depending on: the type of research (e.g., fundamental vs. the creation of a technology), the type of product (e.g. motion detections sensors vs. reconfigurable technology vs. internet bad neighborhoods), the stage of development (e.g. the stage in the design process), and the design team goals/wishes (e.g. advice vs. collaboration). That being said, each collaboration involves value analysis related to the technology in question. This value analysis involves: making intended values explicit, questioning these values, and balancing values with context in mind (examining value trade-offs). To that end, van Wynsberghe proposes here a guideline to address components that pertain specifically to research involving data obtained from online SNSs (Table 1).

It has been suggested that incorporating ethics into ICT research and design be done on a case-by-case basis (Moreno et al., 2008) and we, the authors, would agree with such an assertion. The reason for this being that there are specific details pertaining to each specific research endeavor that may change the ethical dialogue. Added to this is the fact that technologies are and will also be changing, quite rapidly. Consequently, analysis on a case-by-case basis is most important and relevant in order to examine best practices and to report on the success stories (or lack thereof).

This is not to say, however, that it is impossible to suggest a general framework of ethical considerations to be addressed for research involving the use of data obtained from online SNSs. Quite the contrary, there are specific ethical considerations that must be addressed in every

instance; however, the discussion revolving around, and the weight given to, these considerations will differ.

Proposed Guidelines

The discussion of the main ethical issues pertaining to the use of data from online SNSs in this paper represents the first four guidelines presented in Table 1. All of these points may be considered manifestations of the fifth guideline: they are all aspects of the values that the engineer or scientist intends to be realized in their experiments or the resulting product, if they are creating one. Guideline #5 goes beyond a discussion of privacy, however, and demands that the researcher speculate other potential competing values or threats to the intended values.

An important caveat to note here is that van Wynsberghe does not intend this work to be completed by the computer scientist, engineer or designer on their own. Rather, the intention is that this work be completed with the assistance of an ethicist – an individual who is trained to uncover relevant values and to scrutinize these values according to ethical norms, theories and principles. In the case of this work, the computer scientist worked with an ethics adviser to fulfill the value analysis.

1. Make explicit the key actors: direct and indirect subjects, researchers etc.
2. What is the context and what does privacy mean in this context? (location and data content)
3. Type and method of data collection (passive vs active)
4. Intended use of info and amount of info collected
5. Value Analysis: making explicit and scrutinizing intended values of the researchers.

Table 1. Guidelines for best practice when using data from online social networking sites

The question of interest then is how do we put this into practice? Below is a case study based on the work done between an ethics adviser (van Wynsberghe) at the University of Twente, the Netherlands and a computer science student (Been) at the same university[4]. Together we collaborated in order to address the ethical issues related to the use of data obtained from Twitter for the purposes of online identity resolution.

[4] For those interested: Been is still working on his master thesis which might (depending on the graduation) become available online at the University's online repository or can be requested from master-thesis-final@henrybeen.nl

The two researchers met in person on a number of occasions to share information about their respective work. The computer engineer taught the ethicist about his methods and the ethicist taught the engineer about value analysis. Both researchers read information provided by the other to learn in more detail and both researchers wrote about what the other was doing in their own language (i.e. the language of their discpline). In collaboration, they were able to: recognize and make explicit the intended values of the research initiative, to challenge those values and to translate those values into specific design requirements for future research.

Use case: Online identity resolution

Online identity resolution may be defined as matching an individual's online identities with their real world identity and can be seen as a specific variation of regular entity resolution from the domain of databases. The results from such an analysis can then be employed in a plethora of data mining usages for commercial purposes, detection of fraudulent behavior by governments, sociological experiments, research in general and so on.

In this case study, a model for identity resolution was being developed and validated by Been using personal information about an individual from the real world and, in the first step, Twitter. The novelty of this work is that it works directly with Twitter, not with some pre-fetched and cleaned dataset and incorporates mechanics for updating (adjusting) the results over time. The goal was to find the Twitter account of a person, given a piece of real world information. If the model showed favorable results, it would be further developed and tested. This work was done in collaboration with the Dutch ministry of Social Affairs to determine if online identity resolution will be "good enough" to use the results for gathering characteristics about, for example, people who receive social support. This information would then be used to classify welfare receivers as possibly fraudulent or not, using machine learning technologies.

Of particular interest, from the ethics perspective is whether or not one is allowed to use the obtained data for identity resolution analysis in general, as well as for the purposes of fraudulent behavior detection. Of equal interest is what the intentions of the researchers at the University were in creating the resolution identity model and the intensions of the Ministry workers wanting to use the model. With these intentions made explicit the goal is to question whether or not they can, and/or will, be realized in practice. Furthermore, to determine if the model is good enough for further developing and testing, a research experiment was conducted with a set of volunteers who disclosed their personal details and online identities to Been. This experiment by itself may threaten values and the guidelines may be used to separately analyze whether its intentions are realized in practice.

1. Make explicit the key actors: direct and indirect subjects, researchers etc.

Direct users of the system are: for the research experiement, the computer scientist at the University of Twente and, for the intended technology, the data analists at the Ministry that are

hired for discovering social fraud. The direct subjects were those individuals under investigation (whose information was voluntarily given in case of the research experiment) and the indirect subjects were the owners of Twitter and other accounts that were discovered during the resolution and whose data is being stored and analyzed, but that do not belong to any of the original subjects (see Figure 1.)

2. What is the context and what does privacy mean in this context? (location and data content)

The context from which the data is collected is the online SNS known as Twitter. Twitter is an open access social medium in which users post tweets of 140 characters max about anything they want. Tweets are intended to be shared with the world[5], although users can mark their accounts as visible to friends only. As mentioned above, all tweets are currently being archived by the Congress Library and can be openly accessed. Thus, users of twitter are not and should not be under the impression that this information is only accessible to a small cohort of individuals. This does not inply that the tweets can automatically be used for the purposes of detecting fradulent behavior but it does mean that users are not under the false pretense that this information is protected. Note that for other online SNSs, the context and what privacy means in this context can be entirely different.

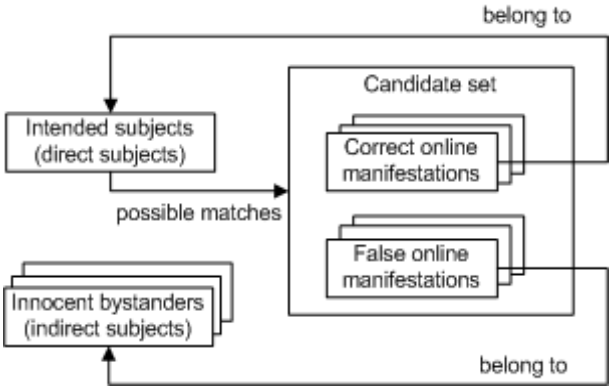

Figure 1: Direct and indirect subjects

3. Type and method of data collection (passive vs active)

For development and validation of the model, data was collected with consent (i.e. actively). Subjects voluntarily provided not only real world information, but also their Twitter information for the validation of the identity resolution model of Been. For the intended technology, subjects provide their real world information for the purpose of obtaining social support. Although they are expected to be aware that this information is used for checking eligability, they are probably not aware that this checking also involves what they say on Twitter.

When applying the prototype in practice, data will be collected without explicit consent from the subjects (i.e. passively). However, before this will be done, collecting such information by investigative authorities will have to become legislated. The main consideration here is expected to be that in order to receive support from the state, you have to allow the state to gather this information to check information provided by the application.

[5] https://twitter.com/tos

4. Intended use of info and amount of info collected

The information collected from individuals in the research experiment was intended to be used for the reliability of the identity resolution model proposed. When deploying the model in practice, the collected data will be used to gather specific characteristics about a person online that are relevant for determining if that person is suspect of fraudulent behavior. Charesterics are, for example, if someone is living alone or not, has undeclared income, or undeclared savings. The following information was collected: all Tweets after signing up for the experiment, the given url (part of a Twitter profile), account name, profile picture and location.

From each tweet, the used language, e-mail addresses and telephone numbers were extracted. If the Tweets had a geotag, these were retrieved to. Further extractions that have been considered, but moved to future work, include: determining location from the tweet text, guessing the subject of the tweet, gathering retweets, gathering followeres and those followed.

Currently, all information collected is used to make a probabilistic approximation of the language, name and location of the account owner. This is then compared to the originally provided details to determine if the account belongs to the given person.

5. Value Analysis: making explicit and scrutinizing intended values of the researchers

The main value intended by the researchers in the creation of the online identity resolution model is reliability i.e., completeness and correctness. These are values that are specific to the model alone and not to a particular application of the model. There are additional and separate values that can be associated with the use of the model. For example, when using the model to detect fraud, the rather abstract value of financial security of the Dutch state is intended to emerge. Also, the values of fairness and justice for citizens at large are intended to increase. This is considered a result of the Ministry's enhanced capability to detect fraud granted from the results of the identity resolution application. Finally, it is important to battle fraud to maintain community support (and taxes) for the social system.

Alternatively, one may suggest that the potential to decrease other values exists. For example, the indirect subjects who owned Twitter accounts but did not belong to the original subjects may feel that their privacy was decreased given that their information was accessed through this approach. This is what is known as a value trade-off, - e.g., the values of justice and fairness for the whole of society may minimize the value of privacy for a subset of that group. Also, one could speculate that false matches might result in individuals wrongfully being classified as **possibly** fraudulent. This is not the case, however, since it is important to note that the Ministry would not, and should not, rely on such methods alone for categorizing behavior as possibly fraudulent or not. Furthermore, the results are only used as a signal that, after verification of the merits of that signal, may lead to an in-depth dossier analysis. In essence, direct and indirect subjects are not considered *suspects* when the technology is used, so the priviliges of the Ministry for lifting privacy restrictions do not apply here.

To adequately address this guideline it must also be made clear how the values relate to the context and amount of information collected as well as the way in which information was collected. Thus, it may also be questioned whether or not the information provided in the context of Twitter be allowed for research purposes in general and furthermore for research purposes of a governmental body. Given that the information is freely accessible in the Congress Library it would be difficult to insist that the information not be used for research purposes – the intention of Twitter and the storage of tweets is for the study of sociological patterns (as opposed to the maintenance of friendships or social networks as in the case of Facebook).

But, does this allow the government access to such information for their own research purposes? To address this question of ethical importance we must ask what the incentives of the researchers are. The incentives of the researchers and the work they are doing is to increase the values of fairness, security, justice and social welfare of the state. This is done in the most effective, reliable and efficient means possible given current research practices. Added to this is the belief that the information found on this site is not subject to the same level of privacy considerations as sensitive personal information or information in the context of one's home. The information used here is from a public site that is freely accessible to the public and not just to a select group of friends or family members. Consequently, in the case of fraudulent behavior detection, the value trade-off identified by these researchers can be justified. In other words, the use of data from online SNSs for this experiment and application falls within ethical limits.

Interestingly, a discussion of online identity resolution for marketing and advertising purposes or for other research purposes like the detection of cyber bully victims to target online assistance would not yield the same results. Using the guidelines would point to differing values at play e.g., economic values in the case of marketing incentives and/or differing subjects involved e.g., a vulnerable demographic requiring greater protection. These factors would shape the value analysis differently and would require further ethical consideration.

Concluding Remarks

The universal and ubiquitous use of computing technologies confronts us with new ethical dilemmas on a daily basis. For example, researchers responsible for the consolidated.db file found on the Apple iPhone, which tracks the geographic location of users without requiring their consent, face the question as to whether this was an intentional threat to privacy or a misunderstanding of the conceptualization of privacy for smartphone users (Cohen 2011). The Facebook T3 project discussed in this paper is another example in which researchers are questioned about their attention to the value of privacy. These are not trivial matters and no clear, pragmatic, guidelines exist to date for how to deal specifically with research involving data obtained from SNSs.

By analyzing work with SNSs on a case-by-case basis (i.e. as a project in its own right but in relation to other research endeavors) according to the guidelines proposed in this paper, it was

possible to identify the ethical limits of the research discussed. Namely, the value trade-off that was identified here was ethically acceptable in light of the research incentives. Specifically, the incentives of the researchers were to increase the values of fairness, security, justice and social welfare of the state. In the pursuit of these values the potential to decrease the value of privacy was identified; however, such a decrease cannot be claimed as such given the context in which the information was obtained (Twitter). Justification of this value trade-off is in contrast to marketing or advertising purposes in which cases the same values would not be intended and thus the same justification could not be made. By making intentions and values explicit, as well as being critical of said intentions and values, it is possible to track the trajectory of values over time and through different contexts, as well as whether or not the intentions and values are being realized once the technology is put into practice.

The purpose of this paper was to stimulate the discussion of best practices concerning the use of data generated and obtained through the use of online SNSs as well as to propose a suggestion for ethical guidelines. In particular, the question of interest for this paper deals with the ethics of using data obtained from online social networking sites for research purposes. This work is considered a kind of value analysis (similar to the approach of value sensitive design of Friedman and Nissenbaum) with the intention of making values explicit in order to scrutinize their ethical justification and whether or not they will in fact be realized when the technology is used. The hope of the authors is to encourage further debate and consideration of ethics in the use of such data.

References

Brey, P. (2000). Disclosive Computer Ethics. Computers & Society, 30, 4, 10.

Cohen, N. (2011). It's Tracking Your Every Move and You May Not Even Know. The New York Times, March 26, 2011. A1.

Friedman, B., Kahn, P. H., & Borning, A. (2006). Human Values, Ethics, and Design. In P. Zhang, & D. Galletta (Eds.), *Human-Computer Interaction and Management Information Systems: Foundations* (pp. 348-372). M.E. Sharpe.

Henderson, T. L. Hutton, and S. McNeilly. Ethics and online social network research – developing best practices. In BCS HCI Workshop on HCI Research in Sensitive Contexts: Ethical Considerations, Birmingham, UK, Sept. 2012. Online at http://www.cs.st-andrews.ac.uk/~tristan/pubs/bcsethics2012.pdf.

Latour, B. (1992). Where Are the Missing Masses? The Sociology of a Few Mundate Artifacts. In W. Bijker, & J. Law (Eds.), *Shaping Technology // Building Society: Studies in Sociotechnical Change* (pp. 225-258). MIT Press.

Manders-Huits, N. (2011). What Values in Design? The Challenge of Incorporating Moral Values into Design. Science and Engineering Ethics 17 (2):271-287.

Moreno, M, A. N. C. Fost, and D. A. Christakis. (2008) Research ethics in the MySpace era. Pediatrics, 121(1):157–161.

Neuhaus, F and T. Webmoor. Agile ethics for massified research and visualization. Information, Communication & Society, 15(1):43–65, 2012. doi:10.1080/1369118X.2011.616519.

Nissenbaum, H (2001) How Computer Systems Embody Values, *IEEE Computer*, March.

Nissenbaum, H.(2011) "A Contextual Approach to Privacy Online," Daedalus 140 (4): 32-48.

Nissenbaum, H. (2004) "Privacy as Contextual Integrity," Washington Law Review 79 (1): 119-158.

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information privacy: Measuring individuals' concerns about organizational practices.MIS Quarterly, 20(2), 167–196.

Raymond, M. (2010). How Tweet It Is!: Library Acquires Entire Twitter Archive. Library of Congress Blog. Retrieved February 25, 2012, from http://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/.

Vieweg, S. The ethics of Twitter research. In Revisiting Research Ethics in the Facebook Era: Challenges in Emerging CSCW Research, Savannah, GA, USA, Feb. 2010.

Van de Poel, I. (forthcoming). Translating values into design requirements. In Philosophy and Engineering: Reflections on Practice, Principles and Process, edited by D. Mitchfelder, N. McCarty and D. E. Goldberg. Dordrecht: Springer.

Van de Poel, I. and Kroes, P. (forthcoming). Can technology embody values? In Moral agency and technical artefacts, edited by P. Kroes and P.-P. Verbeek. Dordrecht: Springer.

van Wynsberghe, A. (2013) Ethicist as Designer: guidelines for the ethicist in the lab. *Forthcoming.*

van Wynsberghe, A. (2012). Designing Robots for Care: Care Centered Value-Sensitive Design. Science and Engineering Ethics, 4.

van Wynsberghe, A. . (2012a). Designing robots with care: Creating an ethical framework for the future design and implementation of care robots. Enschede: University of Twente [Host.

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

Zimmer, M. (2010). Is it Ethical to Harvest Public Twitter Accounts without Consent? MichaelZimmer.org. Retrieved February 25, 2012, from http://michaelzimmer.org/2010/02/12/is-it-ethical-to-harvest-public-twitter-accounts-without-consent/.

Zimmer, M. The Ethics of Twitter Research: A topology of discplines, methods and ethics review boards". Loyola digital Ethics Presentation, Oct 29, 2012. Obtained from: http://www.michaelzimmer.org/2012/10/29/ethics-of-twitter-research-a-topology/