
Ethics and the Long-Term Management of Personal Information

Robert Douglas Ferguson

School of Information Studies
McGill University
Montréal, Québec
H3A 1X1
robert.douglas.ferguson@mail.mcgill.ca

The author(s) retain copyright, but ACM receives an exclusive publication license.

Abstract

Collections of personal information about others have become important resources for researchers, corporations, and governmental agencies. The use of personal information in research and policy raises important questions about ethics and fair use of personal information. In this position statement on the long-term management of personal information, I argue that the amassing of personal information about other people by various groups, especially online, is one of the most important ethical challenges PIM researchers confront today. Using select examples from fieldwork conducted on the research and social networking platform PatientsLikeMe.com, I illustrate how the act of collecting and using personal information about others sometimes violates the expectations and desires for the intended use of that information by those whom the information is about. As a result, it is important that any discussion of long-term management of personal information entails a well-articulated stance on ethics.

Author Keywords

Personal information management; personal health records; personal information; ethics; privacy; Patientslikeme.com; participant-observation

ACM Classification Keywords

K.7.4 Professional Ethics

Introduction

This position statement responds to workshop theme 3 *Managing for legacy*, and explores ethical issues around the intended use, ownership and control over personal information. I reflect on fieldwork data collected in 2009-2010 from the patient social networking and research platform PatientsLikeMe.com¹ and recount an incident that took place that invoked discussion by site users about how they feel about participation and personal health data on the site. The example from PatientsLikeMe.com serves as a cautionary tale for PIM researchers about the importance of critical discussions on personal information along side the field's optimism about innovation.

We arguably live in the age of *big data* in which many of our preferences, behaviors, and interactions online are recorded, analyzed, and compared with data about other people [1]. Over time, those involved in the collection of this data can amass great amounts of information about us without our awareness.

We also live in the age of the *quantified self*, in which we knowingly participate in the tracking of biological, physical, behavioral, and environmental information about to ourselves [2]. Increasingly, we are able to actively record various aspects of our everyday lives and share that information with others. One area in particular in which self-tracking quantifying

¹ Quotes that appear in this paper are from the author's unpublished masters thesis entitled, "*Crowdsourcing health information: An ethnographic exploration of public and private health information on PatientsLikeMe.com*".

technologies have become popular is health and wellness. Various apps, websites, and tools currently exist to help individuals to self-monitor changes in their health or fitness overtime.

At the heart of big data and the quantified self are various senses of *personal information*; information about us that we create when we send or received stuff online, when record and share details about stuff we experience or that is relevant to us [3]. Often times, we own and can therefore control the distribution of our personal information. With the development of seemingly "free" tools and services that help us record, store, and share information, we increasingly share personal information with service providers and other unknown groups in exchange for access to those tools and services.

Personal information can be harnessed to improve tools and services or innovate and develop new ones. These innovations can also have unintended negative consequences for those whose person information is collected as part of its development. For example, in 2012, the New York Times reported how Department store chain Target accidentally a revealed Minneapolis teen's pregnancy to her unknowing family though advertising after successfully developing a pregnancy-prediction algorithm for targeting marketing [4].

Anecdotes like this illustrate why it is important not to overlooking of questions about economic, social, and political consequences of collecting personal data. Boyd and Crawford provide a critical definition of big data as the interplay between technology, analysis, and a mythology, which can be leveraged for both innovation and domination [1]. They argue a critical perspective

entails examining *"which systems driving these practices and which are regulating them"* (p.664) in order reduce seduction by mythologies and shape the direction of innovation. By identifying systems and practices that drive the use, and misuse, of personal information, we can develop a clear stance on ethics that serves as a model for individuals, corporations, and governmental agencies that currently make use of personal data with the ultimate aim of minimizing potential harms to those whose data is collected.

Background on PatientsLikeMe.com

PatientsLikeMe.com (PLM) is a social networking site and data-sharing platform created by brothers Jamie and Ben Heywood with their long-time friend Jeff Cole. The site was originally launched in 2005 as an online community for ALS patients in response to the challenges the family experienced as a result of Jamie and Ben's brother Stephen, diagnosis with amyotrophic lateral sclerosis (ALS) in 1998.

ALS is considered an "orphan disease". Orphan diseases are conditions deemed so rare and affecting so few people that it is difficult to mobilize and recoup costs associated with research and the development of novel treatments for those conditions. In the film 2006 "So Much So Fast" [5] about PLM and Stephen Heywood's life, the founders of the site describe feeling the necessity to become "gorilla scientists" in order to increase the pace of research on ALS. PLM arose out of the need to locate patients around the world and privately fund their own clinical trials with experimental drugs.

Over time, the scope of the site expanded to include people with a variety of different conditions such as,

but not limited to, Multiple Sclerosis, Parkinson's Disease, HIV and Mood Disorders. In 2011, the site expanded its scope again to allow patients with "any condition" to join the site.

Like other social networking platforms, each site user received a profile that they can personalize with their own personal information. What differentiates the profile on PLM from other sites is that it includes a sophisticated Personal Health Record (PHR) that allows users to record changes in self-reported conditions, symptoms, and treatments over time. Data is displayed in the form of visually appealing charts and graphs users. Members of the site can view and comment on the data and can periodically print out their PHR to share with offline health care providers. The site also houses community-centric message boards where users can ask each other questions, share information, and express how illness impacts their everyday lives. The message boards provide a space where site members interact and provide informal social support.

Once members contribute data to the site by using their profile or the message boards, it can only be edited for a short period of time. After that period, members are unable to delete or modify any of the content they submit to the site. Should a member decide to discontinue using the site, they can ask a staff moderator to temporarily deactivate their account and make their profile anonymous but they cannot have their account or data deleted permanently.

Membership on the site is free of cost but participation entailed certain reservations. As per the site's User Agreement, users of the site retained copyright to their data but what appears on the site is considered a copy

with a non-negotiable license to PLM. Data from profiles and message boards are then aggregated, and made available to paying clients, such as pharmaceutical companies. Clients can pay to use premium data collection tools, such as survey and focus groups, to interact more directly with members.

A Critical Incident

During my period of fieldwork in the Mood Disorder community, a critical incident occurred which spurred conversation among its members about the meaning of their data and participation. A data mining bot was discovered on the site, which was later traced to a New York marketing firm. Upon discovery, PLM sent the firm a cease and desist letter requesting that the marketing firm delete all data extracted from the site. Although the firm agreed to stop data mining the site, they were unable to completely remove the scraped data from their database but agreed to quarantine it on their server in order to prevent its future use or distribution.

The story about “the scraping incident” was reported by PLM as well as the Wall Street Journal the latter of which quoted Jamie Heywood stating, “We’re a business, and the reality is that someone came in and stole from us” [6]. Members of the site were quick to notice the way in which PLM made claim of ownership to the data contained on the site, which invoked members to reflect on the nature and value of their participation and data on the site.

Member Reactions to the “Scraping”

Reactions to the scraping incident among site members were mixed. Sometimes members felt that the community, PHR and information on the site positively

contributed to their sense of wellbeing, which they considered a fair exchange for their data,

“I personally don't have time to cruise the internet looking for mentally ill persons on other forums. I find myself happy to have landed here with all these great people to hold my hand. That is payment enough.”

One member, who felt well informed about the site’s business model, saw themselves engaged in a sort of “gift economy” [7] [8] in which data was donated in hope of contributing to the development of cures and scientific advancement,

“I consider my participation in PLM's site, all of it, a donation of sorts. Like giving blood, or time and energy to a political campaign I support, PLM is an organization I believe in and want to give my time, energy and yes, data to. I have spent a fair amount of time making sure I keep my data as detailed as possible for their use.”

Sometimes members stated they knew and consented to their data being used for research but revealed that they were not fully aware of the site’s for-profit business model,

“I sort of knew that PLM collated patient data and used it for research, or shared it with certain companies (eg., pharmaceutical and mental health care industries) but surprised to hear they sell it.”

A member who felt unaware of the site’s business model and did not interpret their data on the site as a donation described feeling somewhat exploited by the site, that profited from the illnesses of members,

"I can't believe I've been here for two-three years .. and I didn't know the information was being SOLD! .. this web-site is monetizing on my misery"

Although PLM was transparent in their User Agreement and other site materials such as their "Openness Policy" about how they use personal information collected on the site, the site violated some members' expectations about compensation and ethical research participation,

"In this day and age, it's not so uncommon and probably not so surprising, but when I take part in Market Research myself I get paid a small amount. Why shouldn't that happen here? We are providing the information that generates the income it is in fact our own intellectual property (unless they've said not in the fine print) I'm thinking I might leave unless we get offered some \$ I'm on a disability pension, life is hard enough ... I can do without knowing that I'm making money for someone without seeing a cent."

As a result of the scraping incident, PLM reported to the WSJ that ultimately 218 members decided to leave the site as a result (6). Because the incident took place near the end of my fieldwork period, I am unable to say what were the long-term impacts on the Mood Disorder community and PLM.

Discussion

The above quotes should illustrate the importance of a proactive ethical stance towards the long-term collection and preservation of personal information, especially for personal information of other people. At the core of research ethics is a concept of human dignity; concern for preserving the autonomy, welfare,

and justice for those who contribute data in knowledge production [9].

To preserve autonomy, it is important that collectors be upfront to those contributing personal information about potential uses, duration of retention, and potential risks of their participation. Effective mechanisms should be in place to ensure consent collect and use personal information is informed and freely given. To ensure welfare, collectors must anticipate potential harms the collection of personal information may cause to individuals and communities. Ideally, those with custodial responsibilities for personal information should prioritize the wellbeing of contributors before their own personal and financial gain. To uphold justice, those who collect personal data must do so in ways that are fair and equitable. They must be sensitive to the ways in which individuals and communities may be more vulnerable to exploitation and harm as a result of their data being collected and avoid instances in which individuals are disempowered through the collection and use of personal information. They should strive to ensure all those who contribute personal information directly benefit to some extent from their contributions.

Conclusion

By exploring an incident that invoked reflection on the how people feel about others using their personal information, the value of taking a proactive ethical stance becomes apparent. The opportunity exists for PIM researchers to develop a more explicit ethical framework for individuals and groups that collect the personal information of others. Critical approaches to PIM help balance optimism about innovations and potential solutions developed by the field.

References

1. Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), 662-679. ACM.
2. Swan, M. (2013). The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data*, 1(2), 85-99.
3. Jones, W. (2010). *Keeping Found Things Found: The Study and Practice of Personal Information Management: The Study and Practice of Personal Information Management*. Morgan Kaufmann.
4. Duhigg, C. (2012, February 16). How Companies Learn Your Secrets. *New York Times Magazine*. Retrieved January 5, 2016.
5. Ascher, S. (Director). (2007). *So much, so fast* [Motion picture]. Balcony Releasing.
6. Angwin, J., & Stecklow, S. (2010, October 12). 'Scrapers' Dig Deep for Data on Web. *Wall Street Journal*. Retrieved January 5, 2016.
7. Mauss, M., & Halls, W. D. (1954). *The gift: Forms and functions of exchange in archaic societies* (No. 378). WW Norton & Company.
8. Malinowski, B. (2002). *Argonauts of the Western Pacific: An account of native enterprise and adventure in the archipelagoes of Melanesian New Guinea*. Routledge.
9. Canadian Institutes of Health Research., Natural Sciences and Engineering Research Council of Canada., & Social Sciences and Humanities Research Council of Canada. (2010). *Tri-council policy statement: Ethical conduct for research involving humans*. Ottawa: Interagency Secretariat on Research Ethics.

Biographic Statement

I am a doctoral candidate at the McGill School of Information Studies in Montréal, Québec, Canada. My Ph.D research explores personal management (PIM) and personal archiving of financial records among young adults. My doctoral research project attempts to answer the following research questions: (1) what records and supporting documents do the personal collections of financial information of young adults typically contain?; (2) how do young adults "manage" their collections of personal financial information within a personal space of information?; and (3) what are the incentives, goals, tasks, and other factors that motivate the management of personal financial information among young adults? My doctoral research is supervised by Prof. France Bouthillier at the McGill School of Information Studies and funded by the Social Science and Humanities Research Council of Canada (SSHRC). I hold Masters and Bachelors of Arts degrees in Social Anthropology from York University in Toronto, Ontario, Canada. As an anthropologist by training, I am interested in individual and group identity construction that occurs through documentation and information-related practices. I am also interested in how human values can be reflected in design of spaces and technology.