

POLICY

Caught in the Web: Informed Consent for Online Health Research

Effy Vayena,^{1*} Anna Mastroianni,² Jeffrey Kahn³

A context-specific approach to informed consent for Web-based health research can facilitate a dynamic research enterprise and maintain the public trust.

There has been an exponential growth in personal health data supplied by users of mobile devices, health apps, and the social Web (social networking sites, online disease support groups, and health-related information sites). These sources and data from tracking of consumers' online behavior coupled with advanced bioinformatics tools offer opportunities for use in health research [for examples, see (1, 2)]. Tweets about disease outbreaks have been correlated with official public health surveillance data and have proven to be an important source of early outbreak detection (3).

A central ethical question is whether individuals who have provided personal information online in nonresearch contexts have consented to research uses. Here, we explore the issue of informed consent for health research performed using information collected from the Web, discuss some limitations of current practices, and offer recommendations for improving consent practices through a more tailored, context-sensitive approach that makes use of the dynamism of the Web-based context. Our proposals are rooted in the ethical imperative of protecting individual rights and respecting autonomy while enabling a dynamic research environment for the advancement of clinical medicine and public health.

A CHANGING CONTEXT FOR INFORMED CONSENT

Health-related research proposals with humans typically undergo prospective review by research ethics committees that ensure that the study is designed and conducted in an ethical manner that protects the privacy and autonomy of individuals through the informed consent process, balances risks

and benefits, and ensures that subject selection is equitable. Informed consent requires that potential participants are provided with adequate information to make an informed and voluntary decision about research participation. Despite the standard practices of obtaining consent, there is a prevalent notion that the process is broken (4). How can an adequate consent process be achieved in health research that involves data collected in the Web environment?

Currently there is limited ethics guidance specifically for research with data collected on the Web (www.aoir.org/reports/ethics2.pdf). Reflected in the dearth of ethics guidelines is either a lack of acknowledgment of this growing area of research or perhaps a sense that this research should not be treated differently from other conventional areas of research. Further, although Web-based research is inherently interdisciplinary because of the range of research areas and diverse sources of data, limited dialogue exists among the represented disciplines in terms of a common set of research principles. New sources of online data and innovative health research applications and the increasingly disparate sources of data challenge traditional approaches to informed consent.

Conventional informed consent models are ill suited because they were not conceived in the context of the evolving applications and functionalities of social media that enable innovative research designs. In addition, traditional approaches to research ethics and informed consent include an ethical distinction between public and private information: The use of publicly available information typically does not require in-

formed consent of the individual, whereas the use of "private" information may require consent depending on whether the information allows an individual to be identified. In an online world, the public-private distinction is increasingly blurred. Should explicit consent be required if a researcher collects and analyzes deidentified Facebook posts that reveal health status or health behaviors? Can health status information shared on social networking sites for patient communities (for example, PatientsLikeMe.com) be used for research without individual informed consent? Is such information properly characterized as public or private?

Current discussions about the public-private dichotomy in the online world include a newer, richer concept that views privacy within "contextual integrity" (5). This approach argues for understanding the importance of the context in which information is located, and determinations of acceptable use are informed by expectations for the use of information within the context in question. This approach places heavier emphasis on the intent of individuals regarding access to personal information rather than on the traditional approaches that demand that researchers protect privacy as a condition of research.

Equally challenging to traditional concepts of informed consent is the control of



Mission control. Data sharing for health research requires an environment of trust.

¹Institute of Biomedical Ethics, University of Zurich, Zurich 8032, Switzerland. ²School of Law and Institute for Public Health Genetics, University of Washington, Seattle, WA 98195, USA. ³Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD 21205, USA. *Corresponding author. E-mail: vayena@ethik.uzh.ch

personal information on the Web. When subjects are traditionally asked to consent to the research use of their data, the limits of that use are spelled out in detail. Personal data posted on or collected by Web sites, however, can be sold or shared and subsequently used in research; thus, it is nearly impossible for users to maintain control of their data, its diffusion, and subsequent uses. As such, the notion of consenting to research use of data loses meaning when the use can involve many unknown researchers and uses in perpetuity. Such open-ended use of data renders the well-established right to withdraw consent to collection and use of personal data for research meaningless. A recent controversial European Union proposal attempts to address the “genie out of the bottle” problem by revising data privacy standards to include a digital “right to be forgotten”—users must be granted an option to delete personal data from the Web permanently (http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf).

DO WEB USERS CONSENT?

Health research using the Web is gaining momentum regardless of available guidance. How informed consent is treated and what it means varies substantially by site and project.

There are two broad types of data gathered through the Web that can be used in health-related research, either by the Web site owner or by third parties: (i) information actively supplied by the user (medical histories, genomic data, and Web posts), and (ii) personal information collected by the Web site while the user is interacting with the site (IP and e-mail addresses, searches, and location data). Both data types may be required by a Web site to enable it to provide the promised services to the user. Many Web services are provided to the user free of charge, while the content that users generate and personal data they provide become trading capital for the companies that provide the services. Web sites often authorize third parties to access their data sets for commercial and research purposes.

The disclosure to users of the potential uses of personal data vary dramatically from site to site. Further, no publicly available studies have yet documented whether users understand or are even aware of the potential uses of their data when they access a site. In reviewing a range of Web sites (6) that collect or contain health data for research, we have identified three general approaches

to consent: (i) research participation as a condition of use of the site, (ii) opt-in to research, and (iii) opt-out of research.

In what we term “condition of use” research participation, Web sites state in their terms of use, terms of service, or privacy statements that they maintain the right to use the data they collect for research, among other uses. By virtue of using the site, the user agrees to research participation. This is equivalent to a so-called browsewrap agreement, whereby the user agrees to the terms of use without any affirmative conduct, such as clicking an “I agree” button (www.eff.org/wp/clicks-bind-ways-users-agree-online-terms-service). It is most likely a carryover from the consumer-oriented sites that use browsewrap approaches as a basic disclosure of policy but without real expectation of careful review and affirmative consent.

The condition of use approach raises three potential concerns. First, the user provides a general consent to a range of uses, including research, rather than consent for a specific research project or research use, and is unable to access the site without giving broad general consent. Second, possible research use is often (but not always) listed among many other uses within the boilerplate language of disclosures and indemnifications, making it questionable that the reader will take notice. Third, the condition of use approach was crafted for consumer agreements to Web site use rather than to accommodate the requirements for informed consent in research. Thus, the condition of use approach stands in stark contrast to the conventional approach to informed consent in health research: a process carefully constructed so that (i) individuals are adequately informed about the project, (ii) the meaning of research participation is clear, (iii) the potential participant makes a voluntary agreement to participate, and (iv) the potential participant is offered the option to withdraw from the research. On this analysis, condition of use research participation does not meet the standards of informed consent except in the most limited and legalistic understanding of consent as agreement evidenced by accepted terms of use.

By contrast, the opt-in approach enables users to agree to participate in a specific research project. Web sites that use opt-in may include a statement with information about the project followed by a link that leads to the project or a requirement to click an “I agree” button to allow research use of personal data. In contrast to condition of use, opt-in par-

ticipation requires an affirmative decision by the user before participating in research. This is equivalent to the common approach to software or other licensed goods in the online world, in which a user must agree to a licensing agreement (also known as a clickwrap agreement) before accessing the product or site. Yet, in that consumer context, research shows that users spend almost no time browsing the text of the agreement before clicking the box, making it unlikely that users opting in to research using this model will carefully read the agreement texts (7). Thus, allowing a research participant the opportunity to review information about the potential research use and confirming participation with an affirmative act comes closer to satisfying the conventional criteria of informed consent, but it is far from clear that the opt-in model achieves the goal of informed voluntary research participation.

With the opt-out approach, Web site users agree to research uses of their data unless they take action to exclude themselves from participation. Users thereby control their data, provided they are aware that the opt-out option exists. For example, some search engines offer users the option of opting out of tracking so that personal data are not collected or stored, although users typically must be sufficiently web savvy to locate the opt-out option. The opt-out approach to informed consent exists in conventional biomedical research when the study poses low risks or when obtaining opt-in consent is impractical and could undermine study design (for example, large-scale epidemiological studies and genome-wide association studies). In these limited contexts, the use of opt-out approaches has been thoroughly debated in the literature and by research ethics committees before being put into practice. Researchers in traditional settings have advocated for wider acceptability of this opt-out approach to informed consent, arguing that it facilitates research while still safeguarding autonomy (www.ieaweb.org/index.php?option=com_content&view=category&id=22&Itemid=54). Whether an opt-out approach to consent in health-related Web-based research satisfies conventional criteria of informed consent requires analysis by consumers, scientists, legal scholars, and research ethics committees.

RECOMMENDATIONS

The health research enterprise continues to evolve along with new possibilities for harnessing online personal health data from

the social Web. Yet, such research is still governed by rules set to address the issues faced in traditional clinical research; this creates a mismatch between the policy requirements for research protections and the types of issues faced in health research using online data. It is important and timely to develop guidance that specifically addresses this new research context. The values that underpin conventional health-research rules remain important. Respect for individual autonomy, balanced and equitable distribution of risks and benefits, and the right to privacy have not lost normative weight because the social Web challenges the approaches we have created to respect them. On the contrary, the challenge presents an opportunity to explore more nuanced and innovative ways of interpreting these values.

It is hard to imagine a useful “one size fits all” consent model for research in the Web environment. Appropriate consent models will depend on the mission of the site, sensitivity and identifiability of the data collected, purpose of the research, and risks and benefits of participation. An interactive process is better suited to meeting the criteria of informed consent. At a minimum, transparent disclosure of the research uses of online personal data are required.

Two types of consent models are well suited for health-related research on the Web. The first is based on the relatively new concept of the Portable Legal Consent (PLC), a legal framework for research consent developed by the Consent to Research project (<http://weconsent.us>; www.inspire-2live.org; www.sagebase.org). It allows participants who are willing to relinquish control of their personal information to attach a one-time research consent to their health and genetic data, which they upload themselves onto the Web site. The data can then be used for research purposes by any researcher who agrees to specific criteria: publication of research results in an open-access forum, no reidentification of participants, and no redistribution of data unless the data recipient agrees to the PLC conditions. The overarching goal of the PLC is to minimize barriers to data sharing and make research data more widely available. Transaction costs related to contacting participants for consent to individual studies are eliminated, privacy concerns are minimized as the data are deidentified to the extent possible, and the participants are informed

of potential privacy risks before consenting. Participants may withdraw their data from the database at any time, but are clearly advised that once data are uploaded, it may not be possible to remove it from all sources (for example, from researchers who have already downloaded, shared, or used the data). The Self Contributed Cohort for Common Genomics Research Study (SCC-CGR) (<http://weconsent.us/about-us>) is the first to implement the PLC.

The PLC improves on existing consent processes in two important ways: (i) the conditions, rules, and restrictions that apply to all research and to researchers who access the data are transparent, and (ii) the risks to participants of sharing their data are clearly articulated. However, in contrast to traditional consent models, the PLC approach, in its proposed state, has some apparent shortcomings: (i) PLC participants must be willing to give up control over their personal health data, including a limited right to withdraw and the choice to decline specific research projects; (ii) deidentifying of data limits its usefulness for some research projects; (iii) PLC relies on a self-selected, well-informed population of computer-savvy users who are unlikely to be representative of the population at large; and (iv) PLC cannot be used for Web data collected for nonresearch purposes.

The shortcomings of the traditional consent process and PLC argue for an approach that is sensitive to the unique aspects of Web-based health research and that harnesses the dynamic aspects of the Web environment. Collaborative and context-specific consent employ the communicative and real-time features of the Web to facilitate a more dynamic approach to informed consent (8). Instead of the traditional approach of a one-time agreement that includes boilerplate text, a user could receive tailored information on research participation with specific choices of options relevant to his or her situation. Collaborative consent might provide a way to address a problem that has challenged even the traditional consent process: how to design effective ways of communicating information to prospective participants. Moreover, transparency requires commitment to clarity and the provision of accurate and appropriate information by researchers. Sites engaged in health research or that allow third parties to use their data for research must modernize communication by making use of the multimedia capabilities afforded by the Web.

PLC will give individuals a way to share (or cede control of) their health-related data. But what about prospective participants who desire greater control? Empirical evidence shows that people care about the way their data are used (9). Giving the user control will also contribute toward building trustworthy relationships and likely increase user participation in research. The “health-information altruist” (10) is willing to contribute to research for the common good. The claim behind the health data rights movement is that individuals should be able to control the sharing of their information (www.HealthDataRights.org). These trends can result in valuable contributions to research only if encouraged by an environment that is conducive to trust.

REFERENCES AND NOTES

1. B. H. Wagenaar, P. S. Sullivan, R. Stephenson, HIV knowledge and associated factors among internet-using men who have sex with men (MSM) in South Africa and the United States. *PLoS ONE* **7**, e32915 (2012).
2. J. Frost, S. Okun, T. Vaughan, J. Heywood, P. Wicks, Patient-reported outcomes as a source of evidence in off-label prescribing: Analysis of data from PatientsLikeMe. *J. Med. Internet Res.* **13**, e6 (2011).
3. R. Chunara, J. R. Andrews, J. S. Brownstein, Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *Am. J. Trop. Med. Hyg.* **86**, 39–45 (2012).
4. E. C. Hayden, Informed consent: A broken contract. *Nature* **486**, 312–314 (2012).
5. H. Nissenbaum, A contextual approach to privacy online. *Daedalus* **140**, 32–48 (2011).
6. E. Vayena, A. Mastroianni, J. Kahn, Ethical issues in health research with novel online sources. *Am. J. Public Health* **102**, 2225–2230 (2012).
7. R. Boehme, S. Köpsell, Trained to accept? A field experiment on consent dialogs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM, New York, 2010) pp. 2403–2406; <http://dl.acm.org/citation.cfm?doid=1753326.1753689>
8. P. Bernal, Collaborative consent: Harnessing the strengths of the internet for consent in the online environment. *Int. Rev. Law Comput. Technol.* **24**, 287–297 (2010).
9. J. Turov, J. King, C. J. Hoofnagle, A. Bleakley, M. Hennessy, Americans reject tailored advertising and three activities that enable it (University of Pennsylvania, Annenberg, 2009); http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1478214
10. I. S. Kohane, R. B. Altman, Health-information altruists—A potentially critical resource. *N. Engl. J. Med.* **353**, 2074–2077 (2005).

Competing interests: The authors declare that they have no competing interests.

10.1126/scitranslmed.3004798

Citation: E. Vayena, A. Mastroianni, J. Kahn, Caught in the Web: Informed consent for online health research. *Sci. Transl. Med.* **5**, 173fs6 (2013).