DOI:10.1145/1743546.1743563

Simson L. Garfinkel and Lorrie Faith Cranor

Viewpoint
Institutional Review Boards and Your Research

A proposal for improving the review procedures for research projects that involve human subjects and their associated identifiable private information.

Researchers in computer science departments throughout the U.S. are violating federal law and their own organization's regulations regarding human subjects research—and in most cases they don't even know it. The violations are generally minor, but the lack of review leaves many universities open to significant sanctions, up to and including the loss of all federal research dollars. The lack of review also means that potentially hazardous research has been performed without adequate review by those trained in human subject protection.

We argue that much computer science research performed with the Internet today involves human subject data and, as such, must be reviewed by Institutional Review Boards—including nearly all research projects involving network monitoring, email, Facebook, other social networking sites and many Web sites with user-generated content. Failure to address this issue now may cause significant problems for computer science in the near future.

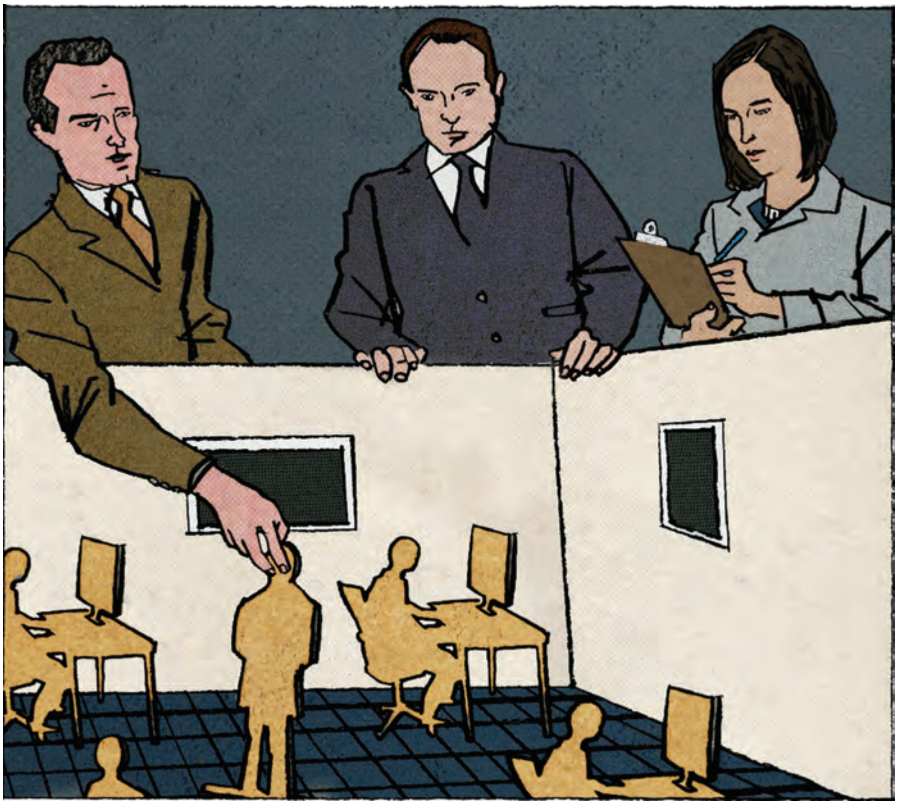

ILLUSTRATION BY DANIEL ZALKUS

Prisons and Syphilis

At issue are the National Research Act (NRA) of 1974[a] and the Common Rule,[b] which together articulate U.S. policy on the Protection of Human Subjects. This policy was created following a series of highly publicized ethical lapses on the part of U.S. scientists performing federally funded research. The most objectionable cases involved human medical experimentation—specifically the Tuskegee Syphilis Experiment, a 40-year long U.S. government project that deliberately withheld syphilis treatment from poor rural black men. Another was the 1971 Stanford Prison Experiment, funded by the U.S. Office of Naval Research, in which students playing the role of prisoners were brutalized by other students playing the roles of guards.

The NRA requires any institution receiving federal funds for scientific research to set up an Institutional Review Board (IRB) to approve any use of humans *before* the research takes

[a] PL 93-348, see http://history.nih.gov/research/downloads/PL93-348.pdf

[b] 45 CFR 46, see http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm

place. The regulation that governs these boards is the Common Rule—"Common" because the same rule was passed in 1991 by each of the 17 federal agencies that fund most scientific research in the U.S.

Computer scientists working in the field of Human-Computer Interaction (HCI) have long been familiar with the Common Rule: any research that involves recruiting volunteers, bringing them into a lab and running them through an experiment obviously involves human subjects. NSF grant applications specifically ask if human subjects will be involved in the research and require that applicants indicate the date IRB approval was obtained.

But a growing amount of research in other areas of computer science also involves human subjects. This research doesn't involve live human beings in the lab, but instead involves network traffic monitoring, email, online surveys, digital information created by humans, photographs of humans that have been posted on the Internet, and human behavior observed via social networking sites.

The Common Rule creates a four-part test that determines whether or not proposed activity must be reviewed by an IRB:

1. The activity must constitute scientific "research," a term that the Rule broadly defines as "a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge."[c]

2. The research must be federally funded.[d]

3. The research must involve human subjects, defined as "a living individual about whom an investigator (whether professional or student) conducting research obtains (1) data through intervention or interaction with the individual, or (2) identifiable private information."[e]

4. The research must not be "exempt" under the regulations.[f]

The exemptions are a kind of safety valve to prevent IRB regulations from becoming utterly unworkable. For computer scientists the relevant exemptions are "research to be conducted on educational practices or with educational tests" (§46.101(b)(1&2)); and research involving "existing data, documents, [and] records…" provided that the data set is either "publicly available" or that the subjects "cannot be identified, directly or through identifiers linked to the subjects''(§46.101(b)(4)). Surveys, interviews, and observations of people in public are generally exempt, provided that identifiable information is not collected, and provided that the information collected, if disclosed, could not "place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation''(§46.101(b)(2)(i&ii)).

Much computer science research performed with the Internet today involves human subject data and, as such, must be reviewed by Institutional Review Boards.

IRBs exist to review proposed research and protect the interests of the human subjects. People *can* participate in dangerous research, but it's important that people are informed, if possible, of the potential risks and benefits—both to themselves and to society at large.

What this means to computer scientists is that *any* federally funded research involving data generated by people that is "identifiable" and not public probably requires approval in advance by your organization's IRB. This includes obvious data sources like network traffic, but it also includes not so obvious sources like software that collects usage statistics and "phones home."

Complicating matters is the fact that the Common Rule allows organizations to add additional requirements. Indeed, many U.S. universities require IRB approval for *any research* involving human subjects, regardless of funding source. Most universities also prohibit researchers from determining if their own research is exempt. Instead, U.S. universities typically require that all research involving human beings be submitted to the school's IRB.

This means a broad swath of "exempt" research involving publicly available information nevertheless requires IRB approval. Performing social network analysis of Wikipedia pages may fall under IRB purview: Wikipedia tracks which users edited which pages, and when those edits were made. Using Flickr pages as a source of JPEGs for analysis may require IRB approval, because Flickr pages frequently have photos of people (identifiable information), and because the EXIF "tags" that many cameras store in JPEG images may contain serial numbers that can be personally identifiable. Analysis of Facebook poses additional problems and may not even qualify as exempt: not only is the information personally identifiable, but it is frequently not public. Instead, Facebook information is typically only available to those who sign up for the service and get invited into the specific user's network.

We have spoken with quite a few researchers who believe the IRB regulations do not apply to them because they are working with "anonymized" data. Ironically, the reverse is probably true: IRB approval is required to be sure the data collection is ethical, that the data is adequately protected prior to anonymization, and that the anonymization is sufficient. Most schools do not allow the experimenters to answer these questions for themselves, because doing so creates an inherent conflict of interest. Many of these researchers were in violation of their school's regulations; some were in violation of federal regulations.

How to Stop Worrying and Love the IRB

Many IRBs are not well equipped to handle the fast-paced and highly technical nature of computer-related research. Basic questions arise, such as,

[c] §46.102 (d)
[d] §46.103 (a)
[e] §46.102 (f)
[f] §46.101 (b)

Are Internet Protocol addresses personally identifiable information? What is "public" and what is not? Is encrypted data secure? Can anonymized data be re-identified? Researchers we have spoken with are occasionally rebuffed by their IRBs—the IRBs insist that no humans are involved in the research—ignoring that regulations also apply to "identifiable private information."

Another mismatch between computer science research and IRBs is timescale. CS research progresses at a much faster pace than research in the biomedical and behavioral fields. In one case we are aware of, an IRB took more than a year to make a decision about a CS application. But even two or three months to make a decision—typical of many IRBs—is too slow for a student in a computer science course who wants to perform a social network analysis as a final project.

For example, one of our studies, which involved observing how members of our university community responded to simulated phishing attacks over a period of several weeks, had to be shortened after being delayed two months by an understaffed IRB. With the delayed start date, part of the study would have taken place over winter break, when few people are on campus. Another study we worked on was delayed three months after an IRB asked university lawyers to review a protocol to determine whether it would violate state wiretap laws.

In another case, researchers at Indiana University worked with their IRB and the school's network security group to send out phishing attacks based on data gleaned from Facebook.[g] Because of the delays associated with the approval process, the phishing messages were sent out at the end of the semester, just before exams, rather than at the beginning of the semester. Many recipients of the email complained vociferously about the timing.

Another reason computer scientists have problems with IRBs is the level of detail the typical IRB application requires. Computer scientists, for the most part, are not trained to carefully plan out an experiment in advance, to figure out which data will be collected, and then to collect the results in a manner that protects the privacy of the data subjects. (Arguably, computer scientists would benefit from better training on experimental design, but that is a different issue.) We have observed that many IRB applications are delayed because of a failure on the part of CS researchers to make these points clear.

It is becoming increasingly easy to collect human subjects data over the Internet that needs to be properly protected to avoid harming subjects.

Finally, many computer scientists are unfamiliar with the IRB process and how it applies to them, and may be reluctant to engage with their IRB after having heard nothing but complaints from colleagues who have had their studies delayed by a slow IRB approval process. While the studies that CS researchers perform are often exempt or extremely low risk, it is becoming increasingly easy to collect human subjects data over the Internet that needs to be properly protected to avoid harming subjects. Likewise, the growing amount of research involving honeypots, botnets, and the behavior of anonymity systems would seem to require IRBs, since the research involves not just software, but humans—both criminals and victims.

The risks to human subjects from computer science research are not always obvious, and the IRB can play an important role in helping computer scientists identify these risks and insure that human subjects are adequately protected. Is there a risk that data collected on computer security incidents could be used by employers to identify underperforming computer security administrators? Is there a risk that anonymized search engine data could be re-identified to reveal what particular individuals are searching for? Can network traffic data collected for research purposes be used to identify copyright violators? Can posts to LiveJournal and Facebook be correlated to learn the identities of children who are frequently left home alone by their parents?

In order to facilitate more rapid IRB review, we recommend the development of a new, streamlined IRB application process. Experimenters would visit a Web site that would serve as a self-serve "IRB kiosk." This site would ask experimenters a series of questions to determine whether their research qualifies as exempt. These questions would also serve to guide experimenters in thinking through whether their research plan adequately protects human subjects. Qualifying experimenters would receive preliminary approval from the kiosk and would be permitted to begin their experiments. IRB representatives would periodically review these self-serve applications and grant final approval if everything was in order.

Such a kiosk is actually permissible under current regulations, provided that the research is exempt. A kiosk could even be used for research that is "expedited" under the Common Rule, since expedited research can be approved by the IRB Chair or by one or more "experienced reviewers."[h] In the case of non-exempt expedited research, the results of the Kiosk would be reviewed by such a reviewer prior to permission being given to the researcher.

Institutional Review Board chairs from many institutions have told us informally that they are looking to computer scientists to come up with a workable solution to the difficulty of applying the Common Rule to computer science. It is also quite clear that if we do not come up with a solution, they will be forced to do so.

[g] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Commun. ACM 50*, 10 (Oct. 2007), 94–100.

[h] §46.110 (b)

Simson L. Garfinkel (slgarfin@nps.edu) is an associate professor at the U.S. Naval Postgraduate School in Monterey, CA.

Lorrie Faith Cranor (lorrie+@cs.cmu.edu) is an associate professor of computer science and engineering and public policy and the director of the CyLab Usable Privacy and Security Laboratory at Carnegie Mellon University in Pittsburgh, PA.