

Supporting Ethical Web Research: A New Research Ethics Review

Anne Bowser
University of Maryland
2117 Hornbake Building, South Wing
College Park, MD 90740
abowser1@umd.edu

Janice Y. Tsai
Microsoft Research
One Microsoft Way
Redmond, WA 98052
janice.tsai@microsoft.com

ABSTRACT

Research ethics is an important and timely topic. In academia, federally regulated Institutional Review Boards (IRBs) protect participants of human subjects research, and offer researchers a mechanism to assess the ethical implications of their work. Industry research labs are not subject to the same requirements, and may lack processes for research ethics review. We describe the creation of a new ethics framework and a research ethics submission system (RESS) within Microsoft Research (MSR). This RESS is customized to the needs of web researchers. We describe our iterative development process, including our assessment of the current state of web research, developing a framework of methods based on a survey of 358 research papers; build and evaluate our system with 14 users to identify the benefits and pitfalls of full deployment; evaluate how our system matches with existing federal regulations; and, suggest next steps for supporting ethical web research.

Categories and Subject Descriptors

K.4.1. [Computer and Society]: Public Policy Issues-Ethics;
H.4.M. [Information Systems Applications]: Miscellaneous

General Terms

Design, Human Factors, Legal Aspects

Keywords

Ethics; Ethics Review; Institutional Review Board (IRB); Research Methods

1. INTRODUCTION

In July 2014, researchers at Facebook published a controversial study exploring how the emotions of Facebook users influence the emotions expressed by their Facebook friends [14]. This research was conducted as an online A/B test, a method considered “the gold standard evaluation for web-based products” by some computer science researchers [9]. Despite the prevalence of this method, public outcry over Facebook’s research was cutting and instantaneous. Some critics took issue with the lack of consent, which those conducting large scale-Internet research studies often find difficult to obtain [15, 17]. Others believed the emotional manipulation may have caused harm. Still others pounced on the realization that neither Facebook researchers nor collaborators at Cornell solicited a formal ethical review [26].

The controversy surrounding Facebook’s study quickly grew beyond the ethics of a single company. The general practice of web

research was called into question as critics articulated a “new privacy truth: legal compliance and sound security practices are not sufficient to meet societal expectations” [12]. In this way, Facebook’s emotional contagion study proved a tipping point, bringing ethical concerns of web research to the mainstream, though web researchers have been grappling with these issues less publically for decades. New technologies, whether videotape in 1995 [16], or public displays in 2013 [15], consistently challenge how existing research methods develop. They also engender new methods, such as data mining, with new ethical considerations.

In academia and at medical institutions, research ethics review is governed by Institutional Review Boards (IRBs). These boards are backed by United States federal policy which dictates that researchers must secure approval for research sponsored by granting agencies such as the National Science Foundation (NSF). In contrast, many industry research labs—especially within the Information Technology sector—are not dependent on federal funding, and, hence, not required to support a formal IRB. Still, as the Facebook example illustrates, the requirement for ethical research crosses sectors. And while IRBs are designed to protect human subjects, they also benefit researchers [21]. They facilitate collaboration by setting common standards within an institution; provide researchers with a framework to critically contemplate the implications of their work; and illustrate the care a researcher takes in their study design to participants, peers, conference chairs, and society as a whole.

We developed a streamlined ethics review program for industry-based research. To do this, we developed a new ethics framework to elicit risk in the specific context of web research. To build this framework, we conducted a survey of methods used in 358 papers presented at 5 top computer science conferences in 2013. This framework was refined and implemented in a digital Research Ethics Submission System (RESS) for use within Microsoft Research (MSR). Our system streamlines the ethics review process with a score calculation and a capacity for automatic approval.

While several companies (including Facebook) have committed to developing a process for internal research review [22], we believe that ours is the first fully implemented system customized for and shared with a community of web researchers.

Specifically, our contribution is:

- A framework for characterizing and evaluating the ethics of the research methods used in web research.
- A digital research ethics submission system customized for the methods used in web research.

We begin by describing standards for ethical human subjects research. Based on these, we explore the current state of academic IRBs. We then describe the methodology used for developing our

ethics framework, provide an overview of the RESS, and report on an early user test with 14 participants. We compare our RESS to federal standards and conclude with limitations and future work: both for MSR, and for the web research community.

2. BACKGROUND

We focus on United States policy because the majority of our researchers work within the US, and because many other nations base their research policies on US law [21]. The guidelines for ethical human subjects research are advanced by the U.S. Department of Health and Human Services (HHS). This agency points to the Belmont Report [2] as a set of overarching ethical guidelines—*respect for persons*, *justice*, and *beneficence*—with applications. HHS also dictates that human subjects research conducted with federal funding must be evaluated by an IRB.

2.1 IRBs in the United States

IRBs assess the use of *informed consent*, an *assessment of risks and benefits*, and the *selection of subjects* to determine whether a research project should be proceed as proposed, proceed with modifications, or be rejected. Notably, not every research proposal must go through full IRB assessment. Federal regulations designate three types of review: exemption from full review, expedited review, and full board review [25]. A research proposal is considered “exempt” when an IRB official determines that the research poses minimal risk to participants, or fits a number of pre-defined parameters. Expedited review is appropriate in situations that involve human subjects *and* minimal risk. Expedited review may be conducted via a single review by the IRB official, or additional reviewers at the official’s discretion. All other proposals require full IRB review by a 5-person board.

2.2 Common critiques

IRBs were created to regulate medical and behavioral science research, and researchers within these fields consider IRBs appropriate and effective safeguards [5]. Others argue that IRBs impose unnecessary and unjustified constraints. Schrag [21] posits that as social scientists have their own shared ethical code, the extension of IRB regulation to social science research constitutes “ethical imperialism.” Schrag also argues that some generic questions posed by IRBs (e.g., does your research involve pregnant women?) are irrelevant outside of these contexts.

Others argue that while broad regulation *is* appropriate, the implementation of IRBs outside of medical and behavioral science is flawed. For example, the idea of “exempt” research, which includes “research involving collation or study of existing data, documents, records, or pathological or diagnostic specimens,” is relatively straightforward in medical and behavioral research [23]. In contrast, “existing data” becomes convoluted in the context of web research (consider: is Twitter an “existing data” set?). Additionally, while research practices involving existing medical data typically poses a low risk to participants in face-to-face studies, research involving existing data in web research may threaten participants through re-identification, or through reuse of data in a new context [5]. Yet, the idea that web research is “exempt” from ethical consideration is supported by many disciplinary norms. Graduates of some traditional computer science programs, taught to study “systems and processes,” may not understand their work as human subjects research [4]. This is a challenge for all ethics review processes, as those most in need of systematic support may be those least likely to seek it.

Additionally, existing IRBs may be a poor fit for web research if reviewers lack relevant expertise. In an early review of online

psychology experiments, Peden [19] observed multiple ethical violations that IRBs failed to notice or correct, such as failing to discuss consent during web surveys. In a follow-up study of 334 IRBs, 62% of respondents reported that their IRBs lack guidelines for reviewing web research, and 74% reported that their IRBs fail to provide training on web research issues [4]. These respondents identified five areas as particularly problematic: privacy, data security, data sensitivity, models for confidentiality and anonymity, and models for consent.

Researchers themselves echo these concerns. Computer security researchers describe how “boilerplate” language such as “no others will have access to the data” may be taken at face value by an IRB without technical expertise, but challenged by a reviewer with domain knowledge [1]. Deciding whether a platform is public vs. private, or assessing the potential harm to re-identification, are similarly problematic issues [4]. Other researchers question: is acceptance of a corporation’s Terms of Service proxy for informed consent [26]? What are reasonable expectations of online privacy [1, 4]?

Securing IRB approval requires that researchers invest their time gathering materials such as protocols and informed consent forms for submission, and also invest time awaiting a decision [23]. In some cases, securing IRB approval can take over a year—a timeline incompatible with the current practices in computer science research, where research is often carried out by students as classroom projects, or by summer interns [9]. Commercial IRBs advertise as time effective solutions to this conundrum, but come at a significant financial cost to the researcher.

Recognizing numerous concerns with the current IRB system, HHS published an advanced notice of proposed rulemaking to improve human subjects research regulations in 2011 [25]. Unfortunately, the proposed changes have neither been implemented, nor adequately address the range of concerns presented above. There have been additional broad efforts to add web research specific principles outside of HHS (e.g., [8]). And, a few researchers within computer science have proposed changes that would directly support the needs of web researchers. Garfinkel and Cranor [6] advocate for a self-serve IRB Kiosk in the form of a website posing a series questions about a proposed research design. This hypothetical kiosk would be able to identify exempt research, or approve expedited, low-risk proposals while helping researchers consider the ethics of their work, but would be unable to contend with proposals designated for full panel review.

In the following sections we introduce our ethics framework and RESS. Our system addresses a number of the concerns described above, by posing domain relevant questions [21], staffing review panels with web researchers [1, 5, 19], and supporting faster processes of submission and evaluation [19, 23].

3. A NEW TYPE OF REVIEW SYSTEM

To introduce ethics review in an environment where it had not existed in a systematic way, and to ensure minimal opposition and resistance, we needed to craft a new framework specific for the scope of research conducted by MSR. We are unable to adopt existing academic IRB system due to the critiques discussed above. Our system must also include an online submission system, and have a fast turnaround time.

3.1 Framework construction

We began by surveying the methodological landscape of web research to understand the most important questions and considerations to build into our RESS. Specifically, we reviewed

the research methods described in 358 publications in 2013 from 5 conferences to create a framework of web research methods. This framework was refined through interviews with 11 researchers. We describe this work, the resulting framework, and key methodological considerations for web research below.

We considered using an existing taxonomy of computer science research, but rejected this idea on two grounds. First, we were unable to find a taxonomy produced since 2006 [11]. Second, we recognized that existing frameworks likely codify existing methodological, and thus, ethical assumptions, which may be problematic. Since our focus is the creation of an internal ethics review, we used the publication history of our research lab to define the research fields pursued (see: limitations).

Using the ACM Digital Library (<http://dl.acm.org/dl.cfm>), we identified the five conferences with the most publications with MSR authors in 2013. These were the SIGCHI Conference on Human Factors in Computing Systems (CHI), the International Conference on World Wide Web (WWW), the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), the ACM International Conference on Information & Knowledge Management (CIKM), and the ACM Conference on Electronic Commerce (EC). We sampled two *full* papers from each *unique* track (e.g., we did not review notes or posters, nor did we sample from a track called “Search I” and also a track called “Search II”). This yielded 358 papers for review, as summarized in Table 1. Note that this sample represents roughly 1.8% of all papers published in the ACM digital library in 2013.

Table 1. Research Contributions Reviewed

Conference	# Full Papers	# Unique Tracks	Sampled Papers
CHI	316	76	182 (58%)
WWW	137	31	62 (45%)
SIGIR	74	14	28 (38%)
CIKM	143	25	50 (40%)
EC	73	17	34 (47%)
Total	743	163	358 (48%)

We read, at a minimum, the abstract and methods of each paper, and recorded the methodologies used. These notes were compiled into a single document and analyzed via an iterative approach inspired by grounded theory [6]. Specifically, we began by conducting open coding to identify key phrases, then arranged key phrases into larger groups, and finally arranged emerging groups into a larger classification scheme, which we refined until satisfied. This process was designed to avoid classifying research based on historical distinctions, such as quantitative vs. qualitative, which may not accurately represent contemporary work.

3.2 Framework validation

We refined and validated our initial framework through semi-structured interviews with 11 MSR researchers in our research institution (again, see: limitations).

3.2.1 Participants

Participants were recruited to represent a range of research interests and geographies. For example, we interviewed researchers with interests including data mining, machine learning, privacy, ICT(4)D, media studies, economics, and socio-digital systems, to name a few. Participants were located in Asia, India, the United Kingdom, and the United States.

3.2.2 Procedure

After describing our research and obtaining consent, we presented each participant with a version of our framework in the form of a flow chart. These charts covered a range of research processes, as presented below. We asked participants questions to “*Think of a research project you are conducting or planning. Please use this decision tree to share your research methodologies with us.*” After walking through the framework and answering questions, participants were thanked and debriefed. Our framework was altered after each of the first 7 reviews. After the 8th review, feedback did not merit additional changes.

3.3 Types of Human Subjects Research

Our framework identifies four general types of research involving human subjects along two axes, *participant awareness* and *environmental manipulation*.

3.3.1 Common types of human subjects research

Our defined research types the following:

- **Type I:** Participants are aware of the research, and researchers manipulate participant environment;
- **Type II:** Participants are aware of the research, **but** researchers **do not** manipulate participant environment;
- **Type III:** Participants are **not** aware of the research, but researchers manipulate participant environment;
- **Type IV:** Researchers collect or obtain personal information about participants **without** participant awareness or environmental manipulation.

Key definitions include the following:

- *Awareness* results from online or real life exchanges between researcher and participant, conducted asynchronously (e.g., through Mechanical Turk) or in real time;
- *Manipulation* characterizes situations in which a stimulus (e.g., an experimental manipulation; a request for data) is designed to produce a specific response. This does not include indirect manipulation, where a stimulus is a facet of the methodology (e.g., the presence of a researcher in a chat room; an observable recording device).

The types are elaborated in Table 2. Note that these research types are not mutually exclusive: any publication may describe research classified in multiple categories. For example, researchers may use data mining to examine the demographics of social media users (Type IV), then go on to recruit a separate set of participants for an interview study (Type II), presenting these as complementary facets of a single research project.

Table 2: Common Types of Web Research with Illustrative Examples

Type	Description	Illustrative Examples	Common methods for data collection
I	Participant awareness: Yes Environmental manipulation: Yes	Lab-based user study, modified data sets (e.g., annotated images) produced upon researcher request	Observation; surveys; interviews; behavioral trace data
II	Participant awareness: Yes Environmental manipulation: No	A diary study of daily habits; a survey on general attitudes towards social media; analysis of existing artifacts shared by participant (e.g., Instagram photos)	Observation; surveys interviews; artifact analysis; diary studies
III	Participant awareness: No Environmental manipulation: Yes	Experiments conducted without participant knowledge (e.g., most online A/B tests)	Online telemetry and usage data collection
IV	Participant awareness: No Environmental manipulation: No	Observation conducted without participant knowledge or consent; data mining studies that collect personal information (e.g., unique identifier; location)	Observation; data mining; artifact analysis

Our research review was designed to let categories naturally emerge from the data. After identifying and refining these categories, we began to critically evaluate how each category is distinct from the others. One important distinction is level of interaction between researcher and participant. While recognizing that this may not be true in every case, as a general rule, Type I research supports the closest interaction between researcher and participant, with levels of interaction progressively decreasing in Type II, Type III, and Type IV research.

Research types also vary regarding locus of control. Type II research often manifests as self-reported attitudes, behaviors, or beliefs. As conversations between a researcher and a participant, these methods generally allow participants a high degree of autonomy regarding how to interpret a particular question or decide exactly what information to disclose. In Type I research, participant autonomy is significant, but may be less strong than in Type II research (consider the degree of freedom afforded by the question “Please tell us about your Facebook use,” typical of a Type II study, versus the question “Please tell us how you would use this perspective Facebook feature,” which might be asked in Type I research). Types III and IV afford less autonomy, as participants are not even aware that research is being conducted. Finally, these categories vary in terms of how research is understood and presented. For example, in Type I and Type II research, authors often refer to *participants* or *subjects* in the *Methods* section of a conference paper. In contrast, Type III and Type IV researchers may describe *data* or *data sets* involved in *experimental manipulations* (referring to both real-time A/B tests and experiments run on archived data). Note that the linguistic conventions of Type III and Type IV researchers are often similar to those of researchers who conduct *experiments* or *modeling* with *real-life* or *synthetic* data in non-human-subjects research. Following constructivist theory, which suggests that human understanding of the world is shaped in part through dialogue, these distinctions are non-trivial.

3.4 The prevalence of different research types

Our classification of conference papers suggested a strong relationship between research type and publication venue. This is natural reflection of each conference’s unique purpose. However, the topical nature of conferences increases the likelihood that researchers will rub shoulders with colleagues of similar epistemological and methodological perspectives, and decreases the likelihood that researchers will be exposed to radically different points of view.

3.4.1 Classification

In light of these considerations we decided to investigate the relationship between research type and venue more fully. Two researchers independently classified 10% of the 358 conference proceedings ($n=36$). Of 43 distinct research studies identified, there were 36 agreements and 7 disagreements (Cohen’s Kappa= 0.79). Disagreements were resolved through closer readings of the papers in question. Note that this process confirms that our classification can be useful for characterizing computer science research to the extent that it is represented in these conferences. We classified the papers in our sample.

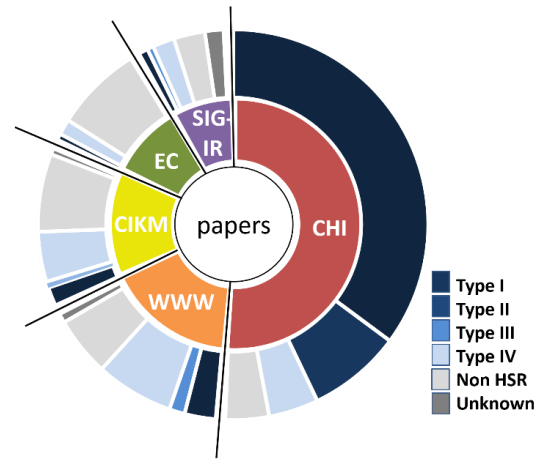


Figure 1: Conference Proceedings Reviewed

Of 358 publications, we identified 263 unique instances of human-subjects research reported in 251 papers (70%). In 96 publications, methods were described in such a way that we concluded these papers did not report human subjects research (27% of total). It was impossible to classify the remaining 3% of our sample. Papers considered non-human-subjects research covered topics ranging from discourse analysis of newspaper articles, to natural language processing of news archives, to evaluation of interfaces without human aid.

Of the 263 unique instances of human subjects research, 157 were considered Type I (60%), 30 were considered Type II (11%), 7 were considered Type III (4%), and 69 were considered Type IV (26%). (Note that these do not add up to 100% due to rounding.) As illustrated by Figure 1, these four types of research were not represented equally among the conferences we surveyed. CHI, for

example, was dominated by Type I research (69% of total publications), and was also the *only* conference to publish research categorized as Type II. In contrast, Type IV and non-human subjects research dominated WWW and CIKM (with these types representing 41 papers, or 82% of total contributions in CIKM; and, 44 papers, or 71% of total contributions in WWW).

3.5 Risk Areas in Research

Using the framework, the evaluation of risk is built into the decision trees stemming from the classification of the research.¹ This section describes key areas for evaluating risk.

3.5.1 Research Settings

Research is conducted in controlled environments, or in naturalistic settings. Some research is conducted in physical environments; other research, online. These environments can be characterized as falling on a spectrum of private to public. For example, a digital ethnography of an exclusive online community is *online research* conducted in a *naturalistic, semi-private* environment.

3.5.2 Recruitment, Enrollment, and Screening

We define recruitment broadly, to refer to both active solicitation of participants and automatic enrollment in online studies. Recruitment may be done in advance of participation, in physical settings (e.g., fliers posted to a discussion board) or online. Note that online recruitment (especially via social media) must be conducted in accordance with site policies. This is an important consideration in both instances where recruitment is distinct from data collection (e.g., a participant is recruited for physical lab study via Facebook) and in instances where recruitment is tied to data collection (e.g., recruitment and data collection are simultaneous for researchers who mine Facebook).

3.5.3 Identifying Participants

While many researchers gather demographic data to support data analysis, we are interested in data gathered for the purpose of understanding whether participants may ethically participate in research. In some research, demographic data are collected for this purpose during recruitment or screening. In other cases, researchers believe that suitability is determined by external partners or platforms (including recruitment agencies, collaborators, and online platforms). In still other cases (including data mining), researchers believe it is impossible to conclusively identify demographic traits.

Generally, research may involve two populations with limited power to consent. Participants with low **capacity** may lack the situational ability to understand the risks and benefits involved in research, and to make an informed decision regarding whether to participate [20]. Participants with low capacity may include those who are socially or economically disadvantaged, e.g., homeless individuals; those with situational sensitivities, e.g., women suffering from abuse; and, those with physical or mental disabilities. Participants with low capacity may consent with the aid of surrogates such as caregivers, and consent from these participants need not be written.

Depending on age, children may also be unable to consent to research. Parental consent may be obtained for participants under 18 (or researchers may request a waiver), and *must* be obtained for participants under 13, as required by US federal law. Note that in some cases research with children conducted in “established or

commonly accepted” educational settings may qualify from IRB exemption [13]. Of course, this begs the question of what constitutes a “commonly accepted” educational setting in the era of online courses, including massively open online courses (MOOCs).

3.5.4 Consent and Debriefing

Determining whether participants have capacity to consent is a separate process from obtaining consent. Participants are often informed about a study prior to data collection. In these cases researchers often (but not always) obtain consent at this point in time. We also identify instances where participants are debriefed, either in addition to or as substitute for information offered prior to participation. Deception plays a role in some computer science research, in the form of misleading information (e.g., regarding the true purpose of a study) or through omission (e.g., being informed that data collection was occurring, but not to what end).

3.5.5 Primary Data Collection for Type I and II

Researchers may document participant actions and behaviors, for example, through audio, video, visual recordings, or notes. In this type of data collection, the researcher is solely and actively responsible for determining what is captured and codified. Data are also collected through automatic recordings of participant behaviors. These recordings may capture data from digital interactions, such as search logs, or from physical interactions, such as biometric data. Type I and Type II research alike rely on participant self-report. The specific mechanism of self-report varies (surveys; interviews; focus groups). But, the type of data that we label as self-report is data provided by a participant about *that participant* (such as self-reported texting behaviors).

Data are also created as participants critically evaluate an existing technology, artifact, or environment. These data are generated in response to and are primarily about *an external environment, artifact, or other entity* (though such data may contain information about participants, as addressed below). Finally, participants may create new data sets. These data may be physical artifacts or data, such as Polaroid photos, or digital artifacts or data, such as digital photographs or image annotations. Note that in this case, data are *created* through intelligent processes rather than *generated*.

3.5.6 Primary Data Collection for Type III and IV

For those conducting Type III and Type IV research, one type of valuable data is generated via *utilitarian use* of artifacts, technologies, and environments. These data may be trace data from physical interactions, such as biometric data collected from sensors, or data from digital interactions. Data sets may be *quantitative*, measuring, for example, the number of books added to a shopping cart. They may also be *qualitative*, e.g., the names of books added to a shopping cart. Type III and Type IV researchers also may be interested in data *produced through intelligent processes* that involve creativity, judgment, or expertise (e.g., movie reviews). Again, these data sets may be quantitative or qualitative. Finally, researchers study data produced by or describing interactions between human beings. Examples of these data include declared relational data (e.g., LinkedIn Connections), data shared between parties (e.g., retweets), and reactions to shared content (e.g., Facebook likes).

3.5.7 Secondary Data Collection

Data about secondary participants may be embedded in data collected from primary participants. “Likes” of a Facebook post

¹ The complete framework can be found at <http://aka.ms/MSREthicsFramework>.

contain the name of that post's author, and A/B experiments analyzing the use of a particular browser may capture data from multiple family members. Additionally, non-sensitive data that are deliberately collected may be intrinsically linked to sensitive data. A data set listing the number of Foursquare check-ins at a *location* of interest also includes data about *humans* who checked in. In some cases, personal or sensitive secondary data can be removed. In other cases, such cleaning would render the data set useless.

3.5.8 Data sharing

Raw data sets are either kept private, or shared with colleagues, with other participants, or with the general public. Similarly, select raw data are shared with researchers outside the project team, with other participants, or with the public at large. Sharing select raw data is different from sharing data in aggregate form. For example, the content of a tweet posted as part of a digital disaster relief campaign is considered select raw data.

Researchers sometimes share raw data containing sensitive information, which we define as data that could damage participants through reconstruction of personal information or through use in a new context. This practice is especially prevalent when a raw data set stripped of sensitive information would lose its value. These key components of web research are codified in our digital research ethics submission system (RESS), described in Section 4.

3.6 Feedback via validation interviews

Interviews conducted with lab researchers allowed us to refine our framework, and helped shape our implementation based on early reactions to the emerging RESS system. On the positive side, multiple interviewees expressed that our process helped them understand the ethics and values inherent in their research. One even mentioned that he planned to follow up on our discussion by talking about ethics with others on his research team.

We recognize that some aspects of our process, such as the categorization of different data types, is unorthodox. Indeed, many of our researchers indicated that they tended to think of their research in terms of “I do data mining,” or “I do ethnographic work.” These conceptualizations are a poor fit for the process we present, and caused minor frustrations. But, it became clear through the interviews that pushing researchers outside of their comfort zones was precisely what allowed them to challenge existing assumptions regarding the ethics associated with certain topics, methodologies, or paradigms.

4. BUILDING A DIGITAL RESEARCH ETHICS REVIEW PROCESS

We use our framework to design the digital research ethics submission system (RESS) which calculates a risk score based on research type and risk area, determining the level of review needed for the research proposal. We provide an overview of the RESS, and then address three workflows in depth: the submission process, preliminary evaluation, and full board review.

4.1 Basic infrastructure

To initiate IRB review in academic contexts, a researcher submits a new project to the designated system. This submission includes a list of researchers, their credentials including certificates of ethics training completion, and a document describing the proposed

research, in detail. IRB reviewers access each proposal, and may review the proposal independently prior to a real-time, virtual meeting. Following each meeting a proposal is approved, rejected pending modifications, or rejected. This process may be repeated until approval is obtained.

Our first departure from this system is our choice of platform. We decided against existing platforms, which can be expensive and time consuming to implement, for various reasons. Unlike their academic counterparts, Microsoft researchers are not subject to federal regulation of their research. Thus, our IRB will either need to be “opt in,” or mandated through corporate policy. For this reason, there is a significant need to establish the value of our RESS by first initiating a small deployment, and then slowly scaling up.

The absence of federal regulations governing industry research also presents an opportunity, enabling us to address the limitations of current IRBs, and to customize our RESS to the needs of web researchers. For our system, we customized an existing conference paper submission system, Microsoft's Conference Management Toolkit (CMT).² Our customized system supports users in three roles:

- *Authors* submit proposals for ethics review by uploading documents (e.g., consent forms and other protocols) and completing a web form. Authors may also use the system to read reviews and download a letter of IRB approval.
- *Reviewers* evaluate research proposals via forms submitted to conference chairs. Reviewers may also use the system to collectively evaluate proposals through a hosted discussion forum.
- *Conference chairs* manage this process by conducting an initial review, assigning proposals to reviewers, moderating discussions, and notifying authors of evaluation outcomes.

4.2 Overview of MSR RESS process

The review process (Figure 2) is initiated when authors submit a research proposal. When prompted by an email notification, chairs conduct an initial review with three potential outcomes: approval through automated expedited review, approval through human expedited review, or designation for full board review (as described below). Authors of automatically approved proposals receive email notification of this decision. For expedited human review, the chair may elect to review the proposal, or to designate a small number of reviewers. Proposals that receive expedited review may be approved without modification, or pending modification. In the later case, authors are prompted to edit their submissions, and receive approval when edits are complete. Proposals designated for full review are sent to a review panel.

MSR has an IRB registered with HHS. All proposals that are approved by the MSR ethics review process receive an official letter of approval attached to their proposal submission, which authors may share with external parties, such as conference reviewers and academic IRBs, at their discretion.

4.3 Author submission

Our submission system is designed as a web form. Each researcher submits a title and abstract for their proposal, and optionally uploads either a consent form, or an application for a waiver of

² The Conference Management Toolkit, <http://cmt.research.microsoft.com>. Interest parties may request an implementation of their own RESS.

informed consent. Researchers then answer a series of questions based on our framework. These questions begin with a general categorization of research as Type I, Type II, Type III, or Type IV. This classification serves two purposes. First, it allows the system to ask users an abbreviated list of questions based on their research type. Second, it allows the system to present users with questions in a vocabulary consistent with their existing mental models.

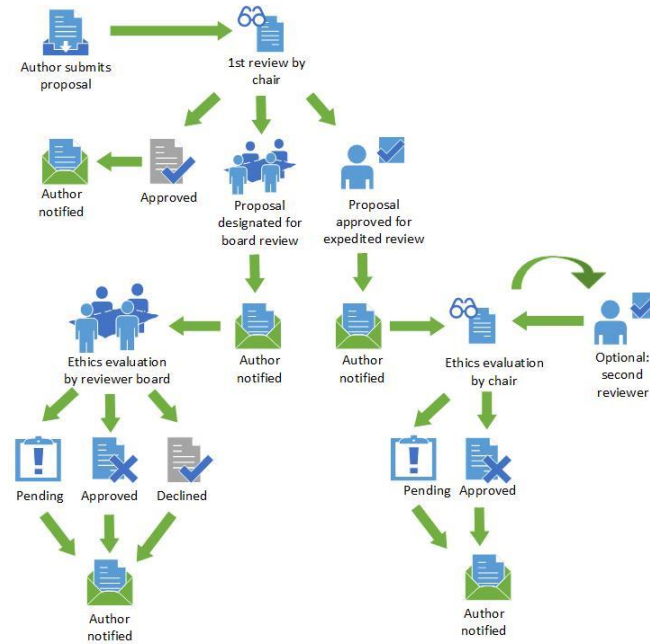


Figure 2. Overview of the research review process

After this initial classification, researchers answer a series of questions using branch logic based on key aspects of the framework described in Section 3. For example:

- For research settings, users are asked whether their research will be conducted in online and/or online settings.
- For recruitment, *enrollment*, and *screening*, users are asked whether they will recruit in advance of participation or during/after participation (e.g., some large display studies).
- Users indicate whether participants may include groups such as children under 13, children under 18, disabled populations, or others with a low capacity to consent.
- Users are asked whether they seek consent prior to research, and/or whether debriefing is utilized after.
- Users describe their primary data collection. For example, Type I and Type II studies may collect data through active documentation (e.g., interviews) or automatic recordings (e.g., biometric data); Type II and Type III studies may collect quantitative or qualitative behavioral trace data.
- For secondary data collection, users are asked whether data about secondary participants may be embedded in data about primary participants, and whether sensitive data is linked to non-sensitive data.
- For data sharing, users indicate whether raw and aggregate data is kept private to the research team, shared with other colleagues, shared with participants, and/or shared publicly.

An initial, automated review follows this submission; if needed, human expedited review or full review follow.

4.4 Initial review

Based on the data submitted by authors, a process of initial review should characterize research as minimal risk or greater than minimal risk. Note that some aspects of a research protocol, such as the use of deception, are risky in themselves. Other aspects, such as the collection of sensitive metadata, are less risky in some cases—e.g., if sensitive metadata is removed prior to analysis—and more risky in others, e.g., if sensitive metadata is collected from children under 13. CMT does not support an easy way to assess these risks, but does support a bulk export to excel. Therefore, we have designed a series of Excel macros to conduct the initial review.

Our process of initial review covers 124 cases where the ethics of a research study may be questionable. Drawing on the linguistic traditions of privacy researchers including [18], we refer to these cases as “flags.” Of the 124 flags, 13 (10%) are associated with an author’s single answer to a question—for example, an answer of “no” to the question “Is research conducted in accordance with site policy?” In addition, 111 flags (90%) are raised when each of two questions are answered in a particular way.

We have designated 3 levels of flags: red, orange, and yellow (Table 3). These flags are derived from three sources: our understanding of canonical ethical guidelines articulated in the Belmont Report, our understanding of ethical questions voiced by those studying web research, and our previous work.

Red flags designate potentially catastrophic aspects of a research protocol. Typically, protocols with red flags are ethically problematic in any environment, but are exacerbated by the nature of web research. Orange flags designate commonly recognized ethical grievances, such as the failure to obtain informed consent in face-to-face research. Yellow flags identify aspects of a proposal that may or may not be ethically problematic in this particular research context, but have been identified as problematic in *some* contexts. The majority of yellow flags are unique to web research, and do not apply to face-to-face contexts.

Table 3: The flagging system for initial review

Flag	Points	Description	Example
Red	15	Egregious ethical consideration; often unique to CS research	Sensitive PII embedded in public data set
Orange	5	Standard ethical consideration	Failure to gain consent in face to face studies
Yellow	1	Possible ethical consideration; likely unique to CS research	Publishing tweets may support re-identification

Each type of flag is associated with a certain number of points. The number of points accrued by each proposal designates the outcome of initial review. Our thresholds are set as:

- *Under 10 points:* Research is minimal risk, and automatically approved. Potential outcomes of review include: accept.
- *10-20 points:* Research is minimal risk, and designated for expedited human review. Potential outcomes of review include: accept; accept with modifications.
- *Over 20 points:* Research is **not** minimal risk, and is designated for full review. Potential outcomes of review include: accept; accept with modifications; reject.

This process is similar to the review via kiosk [9], where an automated system approves select research proposals and designates others for human review. We do believe that the opportunity for initial automated or semi-automated review adds value to a system such as ours, and provide an overview of our implementation for this reason. We also recognize that the implementation itself may be non-ideal, and that each institution seeking to duplicate our RESS (or implement a similar process) will need to carefully construct a framework for initial review suitable to their unique needs.

4.5 Full board review

As described earlier, initial review designates research as automatically approved, suitable for expedited review, or suitable for full board review. We discuss our full review process below.

Each meeting of our RESS is called when a chair selects the “email all reviewers” function, and sends a preloaded template calling the meeting to order. At this point, reviewers are asked to review the proposal submitted by an author, and fill out a form posing the following questions:

- Are the potential risks to participants minimized?
- Are the potential risks to participants reasonable, given the anticipated benefits to participants and/or society?
- Is the selection of subjects equitable?
- Do researchers obtain consent, or provide compelling evidence why consent should be waived?
- Do researchers protect participant safety and confidentiality?
- Please share any additional concerns.

These questions are presented as qualitative response, allowing each reviewer the ability to provide minimal (e.g., “no”) or substantial feedback. In addition, reviewers are asked, “Do you vote to accept or reject the research as currently proposed?” Acceptable answers to this question are “yes” and “no.”

We initially designated each board as “open” for a 48-hour period. This timeframe should allow researchers adequate opportunity to read and review each proposal, while also providing authors with timely feedback. During the 48-hour period, a discussion forum hosted by CMT is activated to allow researchers to discuss each proposal assigned. Following this period, an RESS chair reviews the submitted review forms and online discussions. If the chair observes consensus between reviewers, she decides whether to accept the proposal, reject the proposal, or request modifications. Authors are provided with notification of the RESS outcome, a copy of each review, and a summary of key points made in the discussion forum compiled by the RESS chair. If there is significant disagreement between reviewers, the panel is extended for a second 48-hour period. During this time the chair posts contentious issues to the forum and asks reviewers to discuss the severity of their concerns and brainstorm potential solution. After the second 48-hour period, the chair determines a review outcome by majority vote, and notifies the author.

5. SUBMISSION SYSTEM EVALUATION

We evaluated our proposal submission system through user testing with participants acting as authors ($n=9$) and reviewers ($n=5$). We also evaluated the degree our system adheres to HHS guidelines for the ethical evaluation of human subjects research.

5.1 Author evaluation of interface

Nine participants from our Redmond lab submitted 10 distinct research proposals. Each evaluation began with a meeting where

researchers provided an overview of the review process, offered participants the chance to ask questions, and secured consent. Participants then used our system to submit a proposal of a current or planned project using a think-aloud protocol. Initial review and secondary review (as needed) was conducted through the process described above. Following this assessment, we invited participants to take an online survey describing their experiences with the RESS. This survey contained a mix of open-ended questions, including “What do you consider the value of research ethics evaluation at MSR?” and close-ended questions, such as “Would you submit another proposal for research review?”

Of the 10 proposals submitted, five were approved through automated expedited review; three were approved through human expedited review; and, two were designated for full board review. Though a few usability issues were noted, the majority of authors characterized the system as easy to use. Additionally, most appreciated the use of logic to avoid answering unnecessary questions: the order “*makes sense*.” Participants had mixed reactions to the re-purposing of CMT as a system to support research ethics review. Some complimented the system, and appeared unaware that our RESS was not custom built. But a few objected to the use of legacy language, e.g. the word “abstract.” Our next iteration will address these complaints.

5.2 Reviewer evaluation of interface

In accordance with HHS guidelines, we recruited three colleagues within MSR, one colleague from the legal group within Microsoft, and one external collaborator to act as reviewers. We convened our digital review board for a 48-hour period, asking each of our participants to a) read the two proposals designated for full review; b) submit a web form reviewing each proposal independently; and, c) make at least one post to a community discussion forum for each proposal reviewed (while not all reviewers will be required to post to forums in our implemented RESS, we wanted to evaluate this feature). Reviewers also completed a version of the author survey described above.

Of the five reviewers recruited for our panel, two submitted reviews and three contributed to the discussion forum during the designated 48-hour period. Additionally, two submitted comments and three submitted to the forum within the following 48 hours. This caused us to re-think our timeframe; future panels will unfold over a week, with additional time allocated to discuss contentious issues as needed.

In our user tests, the forums were extremely useful for full board discussion and generated significant and useful feedback for authors. For both proposals, the level of detail voiced through forum comments exceeded the level of detail provided in individual reviews. Regarding usability, reviewers found the RESS interface more difficult to navigate than authors. Some identified issues, such as the use of small and ambiguous icons, are easily corrected. Reviewers also questioned our process, for example asking why feedback was required in both a web form and a forum. This can only be clarified by explaining HHS regulations for ethics reviews in the case of contended proposals.

While we designed an abbreviated RESS to reduce the information burden on authors and reviewers alike, some reviewers wanted more information than our RESS system provided. Specifically, two asked to see consent forms describing the specific risks and benefits of participation (optional for authors), and multiple researchers asked for elaboration on key aspects of protocols. These requests shaped our future implementation, as discussed in Section 6.

5.3 Perceived value of MSR research review

Introducing a corporate research ethics review is a contentious process that may be met with resistance if researchers consider ethics review a bureaucratic hurdle to clear [21]. Only one of our nine researchers failed to find value in our system: *“I believe that most of my projects are pretty straightforward—I put a user in front of a computer and have them click, or deploy it to their workstations; therefore, I’m not too worried about ‘ethics’ issues.”* Still, this researcher contended that the system could have value for those *“sitting at the ethical margins,”* or researchers submitting to conferences requiring a statement of ethical evaluation. In addition to this critique, two of our nine authors questioned why the system was not organized around specific research methods (e.g. online surveys). We address this question in our discussion.

Most participants saw value in implementing a process of research ethics review within MSR. First, participants saw personal value in the system. Some authors indicated that the submission process reminded them of *“forgotten”* best practices, such as debriefing. Others, including those conducting cooperative research with academic institutions, believed that our process could help them submit better applications to academic IRBs. Participants also understood how our RESS could support MSR as a whole. The controversy surrounding Facebook’s emotional contagion study prompted many to comment on the reputational value of research review. For example, one tester considered our system could help *“Avoid an ethical breach (also a public relations disaster) like Facebook recently had.”* Finally, participants saw value in a shared conversation and commitment to addressing ethical issues: *“The value of the review is the conversation it could generate as we all figure out how to do the most ethical work in this challenge space.”*

Most participants familiar with existing IRBs characterized our system as an improvement for evaluating the ethical implications of MSR web research. Testers did share two concerns. One reviewer commented that our RESS is *“a lighter lift but that’s appropriate for the initial program development. Hopefully, as the program evolves, it will become more robust.”* Noting that our process *“feels like filing paperwork (and that’s what usually leads people to misrepresent or miss the ethical concerns that are likely),”* and a second researcher advocated for a dialogical approach supporting conversation between *“fellow researchers.”*

5.4 Compliance with HHS standards

While designing and implementing our system, one key consideration was the degree that our system should comply with HHS regulation for human subjects research. Even if problematic, HHS standards are considered best practices by some researchers.

5.4.1 Requirements for compliance

Federal regulations for the protection of human subjects are outlined in the HHS code of federal regulations under Title 45, Part 46 [25]. This policy applies to all research “supported or otherwise subject to regulation by any federal department or agency” (§46.103). Following a statement of scope and key definitions, HHS §45.46 begins with two conditions of compliance unrelated to IRBs. First, each complying institution must register their IRB, and then file a federal wide assurance (FWA) as a statement of compliance. Second, each institution must implement a statement of ethical principles. Following these conditions, the remainder of §45.46 outlines requirements for IRBs, including requirements for board membership (§46.107) and requirements for securing informed consent or applying for a waiver of consent (§46.116).

5.4.2 Evaluation of MSR RESS compliance

MSR partially complies with the second condition HHS §45.46, relating to registering the board and filing a FWA. Specifically, the MSR RESS is registered with HHS, but we have not submitted a FWA. This decision was made because we have decided to retain the flexibility to define an ethics review process in line with our specific research areas and corporate requirements, and not those mandated by federal process. MSR complies with the second condition of HHS §45.46 by advancing the Belmont Report as a statement of ethical principles. The MSR RESS complies with the vast majority of the remaining HHS requirements. In most cases, compliance is built into the system. For example, §46.115 requires documentation of all proposals and reviews; these are automatically saved by the RESS. In other cases, compliance is set by policy, or through other practices. §46.103 dictates requirements for publishing written procedures describing the review process.

Given our technical infrastructure, there is not a single requirement posed by HHS that **cannot** be met. Thus, it is entirely possible for an industry research lab using our system to fully comply with HHS requirements for the ethical review of human subjects research. Still, we note that some requirements may be problematic for institutions hoping to implement a fully compliant RESS. For example, one of our authors indicated that his research was confidential and should not be shared with reviewers outside of MSR; yet, HHS guidelines dictate that an external reviewer must sit on a full IRB board. Other industry research labs may share this constraint.

6. DISCUSSION

6.1 Key contributions to research and practice

We developed an ethics review system that includes a new framework for characterizing the methods used in web research. This framework serves as the underlying infrastructure for our digital research ethics submission system (RESS). Our framework supports the calculation of risk based on the classification of the research and answers to key questions. We evaluated this framework, and the resulting system, with MSR researchers.

While the majority of our evaluators saw value in the system, a few participants voiced concerns. Two researchers questioned whether research ethics review was truly necessary, given the nature of their work. This concern may be addressed by implementing our system as opt-in. It may also be addressed by designing the system carefully, so that researchers conducting truly low risk research are not required to undergo lengthy evaluation process. Our process of automated expedited review is designed to meet this challenge. Notably, both researchers who failed to see value of our system for their work submitted research proposals that were approved via automated expedited review.

In line with our early interviews, we were not surprised to hear some of our later evaluators advocate for a system that considers research methods as oppose to data types. As one participant who believed she was asked too many questions complained, *“It’s just a survey.”* We respectfully disagree. A survey questioning experience with a prototype of an online discussion forum is not equivalent to a survey collecting information about experience with an existing forum for transgender teens. Our framework recognizes these nuances as classifies these two surveys differently (as Type I and Type II research respectively). Therefore, while we do not consider our system of eliciting data types as opposed to research methods the single best solution for those seeking to evaluate ethics in all contexts, we believe that this system meets our needs by

recognizing the considerations unique to web researchers working at MSR.

6.2 Summary of key tensions

A number of tensions emerged during the implementation and evaluation of our system. First, our system is more effective at supporting the needs of authors, as opposed to reviewers. While we will continue to improve our reviewer interface, we believe that if an imbalance must occur, this is the correct direction: given our process of automated and human expedited review, we expect that a larger number of authors than reviewers will access our system overall.

Second, the intake portion of our system is biased to support complex research proposals over easy ones. However, we believe that asking researchers unnecessary questions is preferable to failing to sufficiently address ethically contentious aspects of a research proposal. We have implemented automated expedited review to balance this bias.

Third, we have designed a relatively conservative system that flags research as potentially problematic from an ethical perspective when, in fact, no significant ethical considerations may exist. Our conservatism is implemented in response to existing standards, which are biased towards classifying research as “exempt” that may, in fact, pose significant risk. We have no desire to stall ethically valid research, and anticipate that any unnecessary conservatism will be corrected via our processes of expedited human review or full board review.

6.3 Limitations and future work

Within MSR, the next step is to fully implement our RESS. We have selected two labs for initial deployment, and will implement full deployment as our system demonstrates the ability to scale. We are also implementing a research ethics training program customized to the needs of web researchers working in an industry setting. We hope that together, these initiatives will effectively support ethical research within MSR.

As we scale up our RESS system we may be able to address some of the concerns raised by our participants. For example, the researcher who advocated for a dialogical review process made a valid suggestion that could help our research community understand and embrace shared ethical best practices. CMT supports the option for reviewers to see evaluations from others; we will support this in future iterations. We may also invite authors to participate in discussions of their research in the online forum. The request for more rigorous review requires careful consideration. On one hand, we deliberately implemented a system that is customized for web research and uses branching logic; adding extra questions or requiring longer answers would mitigate some of the intended improvements to traditional IRB review. On the other hand, we wish to implement a process that is as effective at supporting ethical research as possible.

Finally, we recognize that the effectiveness of any process depends on the dedication of those who participate in it. To this end, our process will continue to prioritize the needs of MSR researchers as it evolves; we expect others who implement ethics review processes in industry settings will follow the same goal. We also note that when researcher needs conflict with established HHS guidelines, difficult decisions must be made.

There are a number of limitations to our work. Perhaps most significantly, our research and development process is biased towards the practices of Microsoft Research. While this is not a limitation for the implementation of our RESS, it is a limitation for the generalizability of our work to other industry and academic contexts. Note that our framework partially address this by

including papers authored by students, academic researchers, and industry researchers alike. Yet, this sample itself is problematic, as we examined papers from only five conferences in the broad field of web research. We invite future researchers to challenge and improve upon our framework by applying it to a greater range of web research in various domains and sub-domains.

To this end, we suggest that supporting ethical web research is truly a community goal. Researchers often define their community based on their research area (e.g. the social science research community; the data mining community). Our classification of the conference papers we reviewed was designed to understand the degree that different types of research are represented in different conferences. We observed that in addition to hosting research around a particular topic, conferences overwhelmingly host research of a certain methodological type. On one hand, researchers within a community are uniquely qualified to evaluate the value of contributions to that community. This is true for research contributions, and may be true for ethical contributions as well [5]. On the other hand, holding discussions within a community of research and practice limits the cross-pollination of ideas. This is particularly significant when the ideas being shared deal with big, out-of-the box challenges—exactly the types of ethical challenges that computer science research provokes. And significant commonalities do exist between researchers conducting similar types of research (e.g., our Type IV research), and between researchers coming from different sub-domains. For example, researchers studying big data and ICT(4)D alike grapple with notions of consent and equity [7, 24].

Researchers agree that considering ethical practices in computer science is not a matter of “hammering offline ethics onto online experimentation” [3]. And, cultural standards should uphold ethical practices as best practices. There are two key barriers to supporting a holistic and shared understanding of ethical best practices within computer science. First, important ethical conversations are occurring within communities of research and practice, and may not be shared across sub-disciplinary divides. Second, the infrastructure that supports academic research will continue to determine how ethics is considered in this space. If industry researchers perceive these processes as outdated—and many do [5]—they will form and follow separate ethical practices.

These silos highlight the need for overarching guidelines that cross sub-disciplines and sectors. Some argue that standards must be truly enforceable, and thus come from a regulatory agency like HHS [24]. We join the call for updated HHS guidelines but also recognize that guidelines at this level are difficult to change, and new challenges are continually posed as technologies evolve. Others suggest that standards should be set or hosted by international associations such as IEEE and ACM [5]. We believe it is particularly important that formal guidelines and standards reflect the actual experiences and expertise of domain researchers, and join this call as well. Guidelines from such associations could be enforced in a number of ways. For example, it could be mandated that conference submissions include a statement of ethical evaluation of the research (ideally, one that is both proactive and retrospective).

ACKNOWLEDGMENTS

We thank Gaoxiang Xu and Vivek Narasayya for customizing CMT, and Carol Boston for reviewing an early version of this paper.

7. REFERENCES

- [1] Aycock, J., Buchanan, E., Dexter, S. and Dittrich, D. 2011. Human subjects, agents, or bots: Current issues in ethics and

- computer security research. In *Financial Cryptography and Data Security*, G. Danezis, S. Dietrich, and K. Sako Eds. Springer Verlag, Berlin, Germany, 138-145.
- [2] The Belmont Report. <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html>.
- [3] Bernstein, M. 2014. The destructive silence of social computing researchers. <https://medium.com/@msbernst/the-destructive-silence-of-social-computing-researchers-9155cdf659>.
- [4] Buchanan, E., Aycok, J., Dexter, S., Dittrich, D. and Hvizdak, E. 2011. Computer science security research and human subjects: Emerging considerations for research ethics boards. *Journal of Empirical Research on Human Research Ethics*, 6, 2 (Jun.2011), 71-83. DOI= 10.1525/jer.2011.6.2.71.
- [5] Buchanan, E. and Ess, C. 2009. Internet research ethics and the institutional review board: Current practices and issues. *SIGCAS Computers and Society*, 39, 5 (Dec. 2009), 43-49.
- [6] Corbin, C. and Strauss, S. 2008. *Basics of Qualitative Research*. Sage Publications, Thousand Oaks, CA.
- [7] Dearden, A. 2012. See no evil? Ethics in interventionist ICTD. In *Proceedings of the 5th International Conference on Information and Communication Technologies and Development* (Atlanta, GA, Mar. 12-15, 2012), ICTD '12. New York: ACM Press, 46-55.
- [8] Ethical decision-making and internet research: Recommendations from the AoIR ethics working committee. <http://www.aoir.org/reports/ethics.html>.
- [9] Garfinkel, S. and Faith Cranor, L. 2010. Viewpoint: Institutional review boards and your research. *Communications of the ACM*, 53, 6 (Jun. 2010), 38-40.
- [10] Gupta, P., Goel, A., Lin, J., Sharma, A., Wang, D. and Zadeh, R. 2013. WTF: The who to follow service at Twitter. In *Proceedings of the 22nd international conference on World Wide Web* (Rio de Janeiro, Brazil, May 13 – 17, 2013). WWW '13. ACM, New York, NY, 505-514.
- [11] Holz, H., Applin, A., Haberman, B., Joyce, D., Purchase, H. and Reed, C. 2006. Research methods in computing: What are they, and how should we teach them? In *Proceedings of ITiCSE 2006* (Bologna, Italy, June 26-28, 2006). iTiCSE '11. ACM Press, New York, NY, 96-114.
- [12] Hughes, J. T. and Tene, O. 2014. The truth is out there: Compliance and security are not enough. *Privacy Perspectives* (3 Oct. 2014). <https://privacyassociation.org/news/a/the-truth-is-out-there-for-big-data-privacy-compliance-and-security-are-not-enough/>
- [13] IRBs and Assurances. <http://hhs.gov/ohrp/assurances/index.html>
- [14] Kramer, A., Guillory, J. and Hancock, J. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *PNAS*, 111, 24 (Jun. 2014), 8788-8790.
- [15] Langheinrich, M., Schmidt, A., Davies, N. and Jose, R. 2013. A practical framework for ethics—the PD-net approach to supporting ethics compliance in public display systems. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays* (Mountain View, California, June 5 - 7, 2013), PerDis '13. ACM, New York, NY, 139-144. DOI= <http://dx.doi.org/10.1145/2491568.2491598>
- [16] Mackay, M. 1995. Ethics, lies, and videotape...In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Denver, Colorado, May 7-11, 1995). CHI '95. ACM, New York, NY, 138-145. DOI= <http://dx.doi.org/10.1145/223904.223922>
- [17] McMillan, D., Morrison, A. and Chalmers, M. 2013. Categorized ethical guidelines for large scale mobile HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France, April 27-May 2, 2013). CHI '13. ACM, New York, NY, 1853-1862. *Proc. CHI 2013*, ACM Press, New York, NY, 1853-1862. DOI= <http://dx.doi.org/10.1145/2470654.2466245>
- [18] Nissenbaum, H. 2009. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, Stanford, CA.
- [19] Peden, B. and Flashinski, D. 2004. Virtual research ethics: A content analysis of surveys and experiments. In *Readings in Virtual Research Ethics: Issues and Controversies*, E. Buchanan Ed. Idea Group, Hershey, PA, 1-26.
- [20] Research involving individuals with questionable capacity to consent: Points to consider. <http://grants.nih.gov/grants/policy/questionablecapacity.html>.
- [21] Schrag, D. 2010. *Ethical Imperialism: Institutional Review Boards and the Social Sciences, 1965-2009*. Johns Hopkins University Press, Baltimore, MD.
- [22] Schroepfer, M. 2014. Research at Facebook. *Facebook Newsroom* (2 October 2014). <http://newsroom.fb.com/news/2014/10/research-at-facebook/>
- [23] Silberman, G. and Kahn, K. 2011. Burdens on research imposed by institutional review boards: The state of the evidence and its implications for regulatory reform. *The Milbank Quarterly*, 89, 3 (Dec. 2011), 599-627.
- [24] Solberg, L. 2012. Regulating human subjects research in the information age: Data mining on social networking sites. *Northern Kentucky Law Review*, 39, 2 (Oct. 2012), 327-358.
- [25] U.S. Department of Health and Human Services 45 CFR 46.
- [26] Verna, I. 2014. Editorial expression of concern: Experimental evidence of massive scale emotion contagion through social networks. *Proc. National Academies of Science USA*, 111, 29, 10779.