Columbia University: Institutional Review Board Annual Educational Conference

Internet Research Ethics

Elizabeth Buchanan, Ph.D. and Maritza Johnson

Endowed Chair in Ethics  Doctoral Candidate
Center for Applied Ethics  Columbia University
University of Wisconsin-Stout  Deprt. of Comp Sc

(Portions of this talk are attributed to collaboration with Laura Odwazny, JD., Senior Counsel, OHRP)
### Evaluating Privacy Settings

**Sharing on Facebook**

These settings control who can see what you share.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Everyone</th>
<th>Friends of Friends</th>
<th>Friends Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your status, photos, and posts</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bio and favorite quotations</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family and relationships</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Photos and videos you're tagged in</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religious and political views</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birthday</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Permission to comment on your posts</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Places you check in to [?]</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact information</td>
<td>●</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- [Customize settings](#) • This is your current setting.
Study Format

- Part 1: Survey of privacy attitudes
- Part 2: Collect sharing intentions
- Part 3a: Identify potential violations
- Part 3b: Present potential violations
Data Collected

- Responses to survey questions
- Sharing intentions
- Identifiers of data items with potential violations (temporary)
- Responses to violations
Data Protection

- Study assigned identifiers
- Did not store the data items in question
- Separate database for identifying information
- Copied data to a local machine
- Secure data transmission? No.
Confounding Relationships

Regulations/Regulatory Boards (Policy)

System/App ideology

Research Participants/Online Norms/
Self-Community Generated
Ethical Frameworks

Researchers/Disciplinary Practices/
Professional Ethics
The Conflicts of Research Ethics 2.0

- Misunderstanding of “human subjects” research vis-à-vis technologies and e-research
  - Rigid regulatory models versus research fluidity (How much wiggle room do we have)
- Intentionality/Context
  - Oversharing…may lead to…
  - Research Creep…may lead to…
  - Mission Creep
- Cultural/Institutional/disciplinary difference
  - (Anglo-American models tend to be more utilitarian-based while Scandinavian countries tend to be more deontological, as they emphasize that the rights of human subjects must never be compromised, no matter the potential benefits.)
Context and Background

- 1999, Frankel and Siang AAAS report
- 2002, AoIR “Ethical Decision Making…
- 2003, Buchanan, Readings in Virtual Research Ethics; Chen and Hall, Online Social Research
- Scattered literature across disciplines
- IRBs facing new lexicon and challenges in their charge to protect human subjects (worms, bots, agents, aggregators)
- A redefinition of what counts as a “human subject” (avatars, turks, etc)
- A redefinition of what counts as “research” (computer science, ie)
- Fed level interest (SACHRP meeting in July 2010, NSF, NIH funding)
Confounding Bases for Review

<table>
<thead>
<tr>
<th>The Same?</th>
<th>New Problems?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethical research is that which seeks to do no harm.</td>
<td>Greased, convergent nature of internet data → harm may be “downstream”</td>
</tr>
<tr>
<td>The greater the vulnerability of the author/participant, the greater the obligation of the researcher to protect the author / subject.</td>
<td>Ubiquitous nature of internet data → who/what/ is vulnerable and when?</td>
</tr>
<tr>
<td>Research integrity itself (good methods and ethics=good research)</td>
<td>Communal nature of internet populations; verifiability of internet agents/subjects; reality of internet data/representation by subjects/participants</td>
</tr>
</tbody>
</table>
New Review Suggestions and Questions

Does the researcher understand the venue or the tool? He/she should be able to articulate a basic level of information about it in his/her protocol.

Is interaction perceived as public or private by the author/participant/subjects?

Does the author/participant consider personal network of connections sensitive information?

How is profile, location, or other personally identifying information used or stored by researcher?

If the content of a subject’s communication were to become known beyond the confines of the venue being studied – would harm likely result?

How do terms of service agreements articulate privacy of content and/or how it is shared with 3rd parties?
New Review Questions

How can researcher ensure that author/participant understands and agrees that content or interaction may be used for research purposes?

Is the data easily searchable and retrievable?

Is the data subject to open data laws or regulations?

What third party policies impact the research?

How long does the third party provider or ISP preserve the data and where?

Can the researcher provide adequate information to participants concerning how the third party will protect their data?

How is protection of privacy of participant/author achieved through anonymization of email content and/or header information?

Regardless of terms of service, what are community or individual norms and/or expectations for privacy?
The Distance Principle as a Guiding Principle

The specificity of Internet research complicates human subjects review in particular.

As the “distance” between the researcher and subject/author/participant decreases, we are more likely to define the research scenario as one that involves “humans.” As the distance increases, we are more likely to define the research scenario as one that does not involve “human.”

The definition of one's data as text versus person (or non-human subjects versus human subjects) may be based on the distance between the product of research and the person who produced it.

The distance principle should be used in tandem with Sveningson’s continuum.
The Distance Principle In Action

- Second Life interview produces data that is near the participant (little distance between researcher and respondent)
- Aggregation of surfing behaviors collected by a bot (greater distance between researcher and respondent)
- Tweeter A (private) $\rightarrow$ followed by Tweeter B (public) $\rightarrow$ Tweeter B retweets A = Tweet A is now visible to Tweeter B’s (any essentially any public) feed (distance implodes!)
Regulatory Exemptions

- **101 (b)(1):** Online environments have become "normal" educational practice and as such are appropriate for exemption. Two particular issues were addressed: Is the particular intervention typical in that learning group? And, was the intervention ongoing before a researcher studied it?

- **101 (b)(2):** This exemption does not apply to observations of minors, and thus, robust age verification measures must be taken by the researcher. This could include age verification software, proxy for age verification, or statements of age verification embedded throughout a given research instrument or venue. From the regulatory perspective, check boxes of verification may be sufficient.
Regulatory Exemptions

- 101(b)(2) and 101(b)(4): Regarding online observation and exemptions for existing data, the issue of researcher "entry requirements" and "threshold of access" is significant. The OHRP standard is "what is readily ascertainable." If there are costs, restrictions, or other measures that prevent data from being "readily ascertainable," the exemption does not apply and the data should be considered private.

- 101(b)(2) calls attention to identifiers. OHRP has not issued a formal statement on IP addresses as PII, however, for purposes of the HIPAA Privacy Rule, one office of HHS (the Office on Civil Rights) has opined that an IP address is considered to be a direct identifier of an individual. Other European data regulations do consider IPs as identifiers.
102(f): New forms of representations, such as avatars, are considered human subjects if PII about living individuals is obtained. PII can be obtained by researchers through scraping data sources, profiles or avatars, or other pieces of data made available by the human "behind the avatar." Observation (per 101(b)(2) and 101(b)(4)) may be an acceptable exemption for intervention or interaction.

102(i): Minimal risk: Internet research presents many different possibilities for reconceptualizing minimal risk. The concepts of universal precautions, individual precautions and responsibility are key. Researchers and boards must balance presenting risks related to the specific research with risks related to the technologies in use. Sample language could be included in information sheets and consent documents, such as "There are potential risks of data loss, data manipulation, or unauthorized access by outside parties in this form of research. All appropriate precautions will be taken to ensure the security of the data."
Regulatory Exemptions

- Additional risks associated with online research include that there is often no direct contact with participants, so distress may not be evident to the researcher; breaches of confidentiality; failure of measures to control invalid access.

- 103(a): According to regulations, an institution or agency that only releases data is not "engaged" in HS research. This has implications for data management banks and repositories. There is an exception: if the institution is the direct awardee of an HHS grant for the research, the institution releasing data would be presumed engaged on that basis.
Regulatory Exemptions: IC

- 116 (d): Elements of informed consent: Many IRBs report waiving elements of IC for minimal risk research. This should be consistent in the Internet research setting as well. Best practices should include "knowledge checks" throughout the IC process to ensure understanding, especially in Internet research settings.

- In some online environments, obtaining consent is impracticable. As with onground research, if the scientific value overrides the need for IC and the criteria for waiver of consent appear to be met, the researcher should clearly justify these conditions.

- 117: Documentation of IC will depend on jurisdiction but OHRP policy holds that electronic signatures are valid if allowable under the pertinent law.
Best Data Management and Security Practices in a “Cloudy” Environment

- Grid and cloud research/Third party apps demand:
  A data management plan that correlates with a tiered data security approach
  A plan of shared responsibility between and among the researcher, information technology departments/CIOs, and the IRB/HREC. A plan for the data custodianship should be established within this model of shared responsibility.
  The research review board should be involved early in the development of any data management plan.
  A data "destruction" plan is necessary. Given the emergence of cloud storage and management, institutions must be clear on their "data hygiene."
  107(a): OHRP strongly supports supplementing boards with expert consultants.
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>A popular online social networking site, designed originally to mimic a school yearbook.</td>
<td><a href="http://www.facebook.com">http://www.facebook.com</a></td>
</tr>
<tr>
<td>Flickr</td>
<td>A social networking site built around sharing photographs.</td>
<td><a href="http://www.flickr.com">http://www.flickr.com</a></td>
</tr>
<tr>
<td>Livejournal</td>
<td>A blog hosting site that hosts both individual and community blogs.</td>
<td><a href="http://www.livejournal.com">http://www.livejournal.com</a></td>
</tr>
<tr>
<td>MMORPG</td>
<td>An acronym for &quot;Massively MultiPlayer Online Role Playing Game&quot;. These types of games often involve hundreds of players participating in virtual worlds.</td>
<td>The &quot;World of Warcraft&quot; is a popular MMORPG.</td>
</tr>
<tr>
<td>MySpace</td>
<td>A popular social networking site.</td>
<td><a href="http://www.myspace.com">http://www.myspace.com</a></td>
</tr>
<tr>
<td>Second Life</td>
<td>An online virtual environment, run by LindenLabs.</td>
<td>The Second Life homepage, which includes screenshots of the virtual environment can be found at: <a href="http://www.secondlife.com">http://www.secondlife.com</a></td>
</tr>
<tr>
<td>Tumblr</td>
<td>A blog hosting site that emphasizes the sharing of multimedia content.</td>
<td><a href="http://www.tumblr.com">http://www.tumblr.com</a></td>
</tr>
<tr>
<td>tweet</td>
<td>A term for a single post on Twitter. A blogging site that limits messages to 140 characters.</td>
<td>An example of a &quot;tweet&quot; from the NASA Twitter page can be found at: <a href="http://twitter.com/NASA/status/20146648791">http://twitter.com/NASA/status/20146648791</a></td>
</tr>
<tr>
<td>Twitter</td>
<td>A blogging site that limits messages to 140 characters.</td>
<td><a href="http://www.twitter.com">http://www.twitter.com</a></td>
</tr>
<tr>
<td>Wordpress</td>
<td>A blogging tool and online publishing platform.</td>
<td><a href="http://www.wordpress.com">http://www.wordpress.com</a></td>
</tr>
<tr>
<td>World of Warcraft</td>
<td>A popular massive multiplayer online role playing game centered around completing tasks within the fantasy world.</td>
<td>The World of Warcraft homepage, which includes screenshots of the virtual environment can be found at: <a href="http://www.worldofwarcraft.com/index.xml">http://www.worldofwarcraft.com/index.xml</a></td>
</tr>
<tr>
<td>Youtube</td>
<td>A large video sharing website on the Internet, owned by Google.</td>
<td><a href="http://www.youtube.com">http://www.youtube.com</a></td>
</tr>
</tbody>
</table>
API

- Application programming interface
- Facebook, Myspace, Twitter, Flickr, iPhone apps
Facebook API
Best Practices?

- APIs
- Online services
The New York Times is requesting permission to do the following:

- Access my basic information
  Includes name, profile picture, gender, networks, user ID, list of friends, and any other information I've shared with everyone.

- Send me email
  The New York Times may email me directly at

Access my data any time
The New York Times may access my data when I'm not using the application.

Access my profile information
Likes, Music, TV, Movies, Books, Quotes, About Me, Interests, Groups, Birthday, Hometown, Current City, Website, Education History and Work History

Use of this data is subject to the The New York Times Privacy Policy · Report App

Logged in as Barbara Krasnoff (Not You?)

[Buttons: Allow, Don't Allow]
Thank you!

(Funding for this research comes from the National Science Foundation; The views and opinions represented here do not reflect the opinions of the NSF.

Questions/Follow up: elizabeth@internetresearchethics.org or buchanane@uwstout.edu

maritzaj@cs.columbia.edu