

# Conducting Research on the Internet: Emerging Ethical, Regulatory and Practical Considerations

Elizabeth Buchanan  
Associate Professor  
University of Wisconsin-Milwaukee

Endowed Chair in Ethics  
University of Wisconsin-Stout

# Research subjects aren't what they use to be...



## Speaker:

**Dr. Elizabeth Buchanan**  
Associate Professor  
School of Information Studies  
University of Wisconsin, Milwaukee

Date: Friday, September 24, 2010

Time: 12 noon—1:30 pm

### Location:

University of Pittsburgh  
Oakland Campus  
Frick Fine Arts Building  
Auditorium (Room 125)

### ON-LINE REGISTRATION:

[www.urb.pitt.edu/crj.aspx](http://www.urb.pitt.edu/crj.aspx)

### NO REGISTRATION FEE

Please join us for a reception with speakers immediately following the

## Conducting Research on the Internet: Emerging Ethical, Regulatory, and Practical Considerations

Dr. Elizabeth Buchanan will provide an overview of some of the more controversial internet research ethics dilemmas that face investigators and IRBs. Taking a case-based approach, she will discuss various ways of considering these issues. She will review internet-based subject recruitment and survey research activities as well as experimental or observation studies where researchers systematically manipulate online avatars and observe virtual reactions, or actively participate in chat rooms and collect data on others' responses. Ethical, regulatory and pragmatic challenges associated with electronic transmission and storage of research data will also be examined.

# Special thanks

- Patty Orndoff
- Chris Ryan
- Pitt IRB
- Pitt Community

# Quick Grounding Points

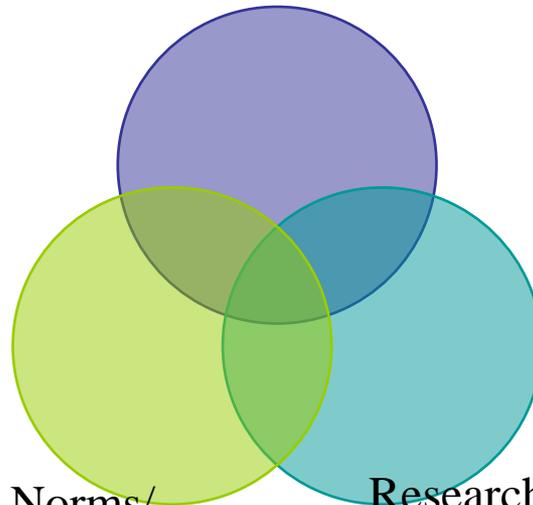
- IRB member (social science and medical school)
- 1998, dissertation, using all online data collection methods
- Online social science researcher
- NSF funded projects
  - Internet Research Ethics: Discourse, Inquiry, Policy
  - [InternetResearchEthics.org](http://InternetResearchEthics.org)

# Context and Background

- 1999, Frankel and Siang AAAS report
- 2002, AoIR “Ethical Decision Making...” (Watch for 2010 revision!)
- 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
- A redefinition of what counts as a “human subject” (avatars, turks, etc)
- Fed level interest (SACHRP meeting in July 2010, NSF, NIH funding)

# Confounding Relationships

Regulations/Regulatory Boards (Policy)



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

Researchers/Disciplinary Practices/  
Professional Ethics

# Confounding Bases for Review

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

# For Example...

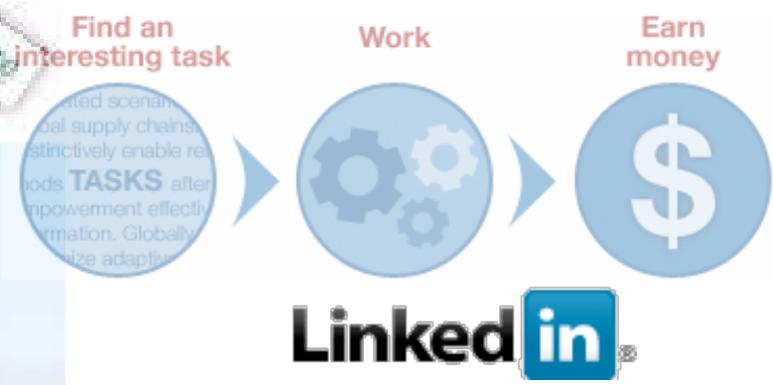
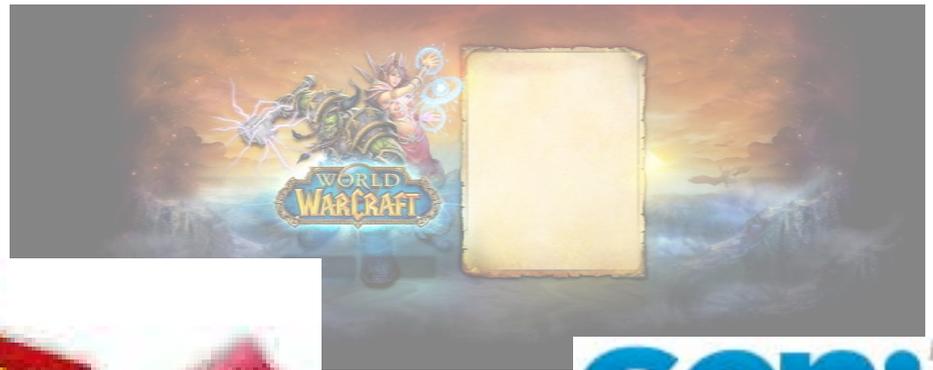
- A researcher is using the website "Gay Bombay" to study gay Indian men's attitudes, and the IRB was worried that since homosexuality is “illegal” in India, would participation get the respondents in trouble somehow?
- Banked data? Genetic, textual, video, audio? (ICPSR, UK Qualidata)
  - New study, JERHRE, September 2010, on reconsenting and genetic research
  - New NSF data sharing initiative (Press Release 10-077: Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans...are you ready?)
- A researcher wanted to use a public list archive, but—in order to post, membership was required. Must he gain consent? Does this fit the “public park” analogy?

# For Example...

- Online survey tools—who owns the data? where is the data? (Buchanan and Hvizdak, 2009)
- Cloud computing— “[C]loud services... typically result in data being present in unencrypted form on a machine owned and operated by a different organization from the data owner”
- 136.159.7.150 - - [09/Sep/2010:09:16:51 -0600] "GET /worm HTTP/1.0" 404 574 "-" "Chrome/1.414.213" (Whaaa?) (Buchanan and Aycok)
- Can a researcher use mechanical turks to complete research related tasks, eg, survey responses, without IRB oversight?
- Is aggregated facebook data really anonymous? (Zimmer, 2009/2010)
- Interesting reinterpretation: [Virtual Milgram](#) (Slater, et al)

# An Array of Issues Awaits....





# Anonymity/ Confidentiality

- Distinction between anonymous and confidential
- *“Our first duty as researchers is to honor the promise of confidentiality”* (Easter, Davis, & Henderson, 2004).
- Is there a truly secure online interaction? What type of Internet location/medium is *safest*? Is an “anonymous” survey possible? How will subjects/participants be protected? Is encryption enough?
- Can one be anonymous online? One may have a “different” identity (e-betty), but that is still “me.” If e-betty is portrayed in research on an electronic support group for a medical condition, will she be identifiable? If so, at what risk?

# Revealing Identities

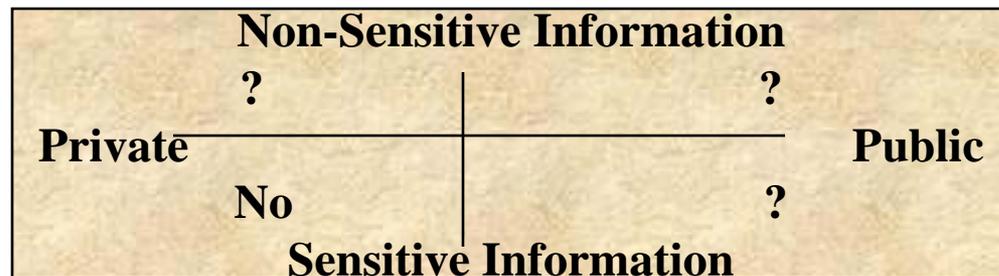
- How should online participants be identified in research reports given the traceability and trackbackability of online data?
- Screen names, pseudonyms of screen names?
- By changing screen names, do you detract from the “reality” or “reputation” of the participant?
- Text searches can reveal more context than a researcher may in her reporting (potential risk)
- Allow participants to make this decision? Part of informed consent?

# Possibilities for Consent: Lawson, 2004

1. consent to having their nickname and communicative text used for data analysis only (no publication of name or text);
2. consent to having either their nickname or text published in an academic work, but never together (i.e., no identifiers);
3. consent to having either their nickname or text published in an academic work, but never together (i.e., no identifiers) and providing they get to see the 'write up' prior to publication;
4. consent to having both their nickname and text published in academic work, thereby being credited as the authors of their own words; or
5. consent to having both their nickname and text published in academic work, thereby being credited as the authors of their own words, providing they get to see the 'write up' prior to publication (p. 93).

# Public and Private Spaces

- Is a particular forum, listserve, chat room, bulletin board, etc considered *by its members* to be a public space or a private space? What expectations of privacy exist? Members only? Public newsgroups? What sort of methods are being used? This might not be reviewable, eg, discourse analysis?
- What role does the researcher play in the space? (Observer, participant, member, other?)
- What is the content of the data? (Sensitive/non-sensitive?) Medical information, eg, Caring Bridge?
- Sveningsson's continuum:



# Ownership/Stewardship of Data?

- With f2f research, the researcher, eg, conducts an observation. Writes field notes. Returns a report to the participant when completed. Owns “it.”

Versus:

- A researcher conducts an observation of some newsgroup interactions. A log/transcript is generated. Researcher has a copy. So do the participants. So does the server/administrator of the news group. Who owns “it?”
- How long does e-data last? (“I will destroy the data in 5 years...” may mean nothing in an online context where researchers are not in control)
- Advise researchers to inform participants/subjects about the longevity and potential risk of data intrusion.
- Cloud computing
- Data banking/repositing

# Respect for Persons/Autonomy

- Practical challenges in obtaining informed consent (fluidity, changes in group membership, etc)
- Verification of understanding (the cornerstone of informed consent is understanding!)?
- Will something from a public space come back to haunt us?
- Use of archived quotes never intended to be represented in research?
- Click boxes? Hard copies? Blanket statements necessary (*“I understand that online communications may be at greater risk for hacking, intrusions, and other violations. Despite these possibilities, I consent to participate”*)?
- Blogs (don’t meet the definition of human subject as in 45cfr46.102f)
- Are avatars “human subjects”?

# Recruitment

- Equity/fair representation in subject pool may not be possible: in our protocols, we may see “unjust” subject selection based on type of site— eg, WoW
- How does the researcher enter the research space to begin recruiting? Site owners, moderators, gatekeepers are key in some fora
- What if some in a community consent, others do not? Researchers should have plans for this reality.

Ok, Breathe....



# 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.

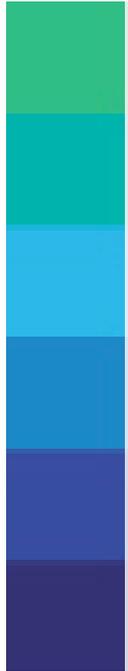
As the “distance” between the researcher and subject/author/participant **decreases**, we are more likely to define the research scenario as one that involves “humans.” As the distance **increases**, we are more likely to define the research scenario as one that does not involve “human.”

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) → followed by Tweeter B (public) → Tweeter B retweets A = Tweet A is now visible to Tweeter B's (any essentially any public) feed (distance implodes!)



Have a **question?** Want to **learn more?**

visit the

INTERNET RESEARCH ETHICS

**COMMONS**

What Didn't I Cover?

# Thank you!

- Email: [elizabeth@internetresearchethics.org](mailto:elizabeth@internetresearchethics.org)

