The New Normal

Elizabeth A. Buchanan, Ph.D.
University of Wisconsin-Stout
buchanane@uwstout.edu

Disclosures

- None related to this educational activity.

Objectives

- Consider long-standing and current issues in Internet research (broadly defined)
- Recognize how we studied research ethics in this domain and how guidelines/best practices emerged from our research
- Review how your these discussions related to your MRPs and future research
- Consider IRB review and considerations

Internet Research, circa 2004

- Internet-based research, broadly defined, is research which utilizes the Internet to collect information through an online tool, such as an online survey; studies about how people use the Internet, e.g., through collecting data and/or examining activities in or on any online environments; and/or, uses of online datasets, databases, databanks, repositories.
  - Internet as a TOOL FOR research or…
  - Internet as a MEDIUM/LOCALE OF research
- TOOL=search engines, databases, catalogs, etc…
- MEDIUM/LOCALE=chat rooms, MUDs, MOOs, newsgroups, home pages, MMORPGs, blogs, skype, tweeting, online course software, etc
- Increasingly, the line between tool and locale is blurring in the realms of social media, mobile/cell, computer-based research.

Early Internet Research Considerations

- Public/Private
- Privacy, Confidentiality, Anonymity
- Recruitment and entering spaces
- Vulnerable populations
- Data Security
- Consent
- Minors
- International research

    - How were IRBs reviewing research and on what bases?
    - And, are these new ethical issues or just “different”?

Early Cases (circa 2002-2005)

We had a researcher using the website "Gay Bombay" to study gay Indian men's attitudes, and the board was worried that since homosexuality is illegal in India, would participation get the respondents in trouble somehow? (Transborder issues, risk, consent)

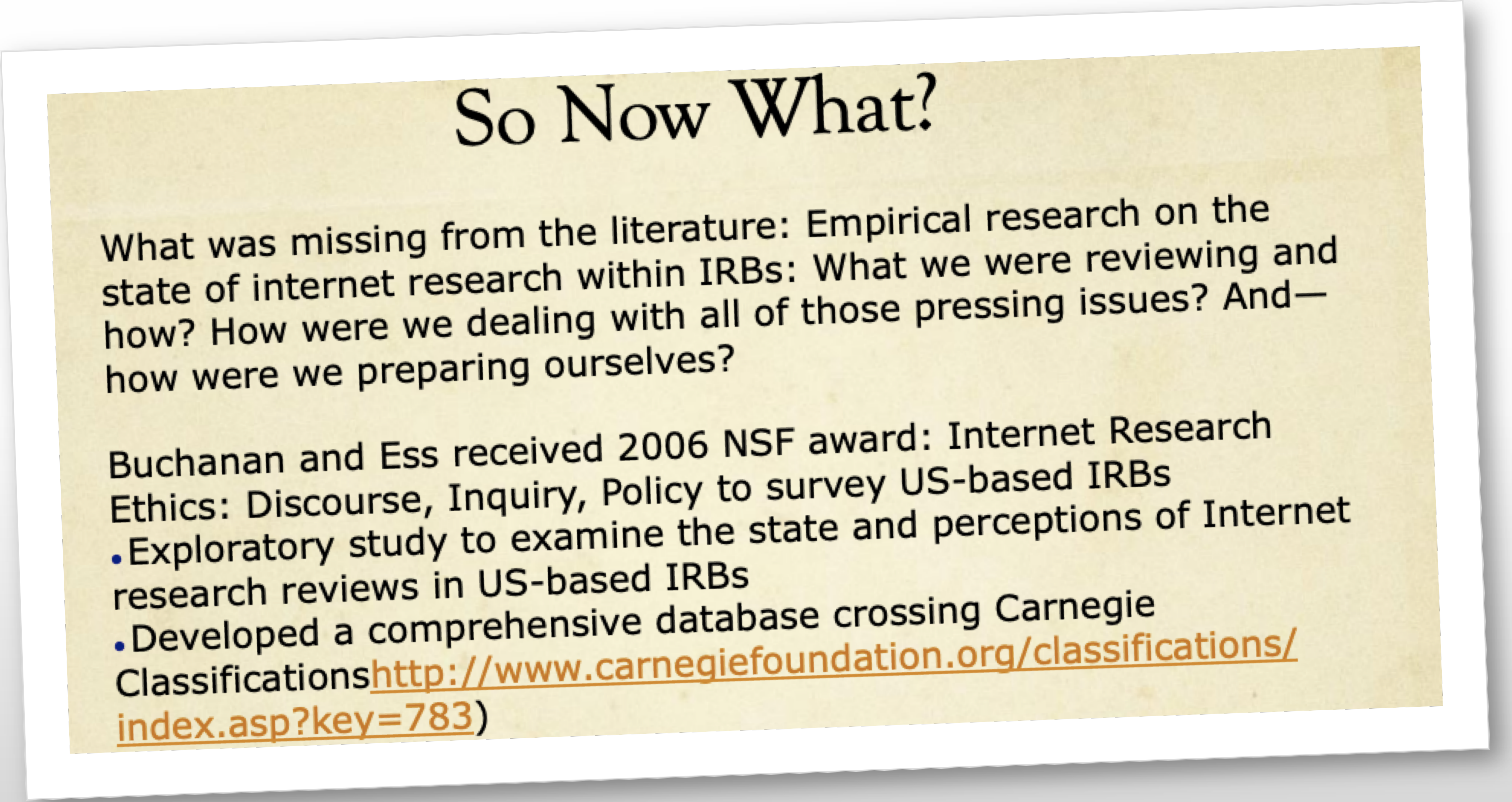
A male psychologist posed as a disabled woman to study the relationships developed in an online group. He did not disclose his researcher status.

A student wishes to analyze blog postings as part of her Master’s thesis. Must she seek IRB review? If she does not, might she face journals who will not publish her work because it was not approved?

A researcher wanted to use a public list archive, but—in order to post, membership was required. Must he gain consent? (No longer fits the “public park” analogy?)

Can a researcher use mechanical turk (to complete research related tasks, eg, survey responses) without IRB oversight?

What Resonates With You--Now?



So Now What?

What was missing from the literature: Empirical research on the state of internet research within IRBs: What we were reviewing and how? How were we dealing with all of those pressing issues? And—how were we preparing ourselves?

Buchanan and Ess received 2006 NSF award: Internet Research Ethics: Discourse, Inquiry, Policy to survey US-based IRBs

- Exploratory study to examine the state and perceptions of Internet research reviews in US-based IRBs
- Developed a comprehensive database crossing Carnegie Classificationshttp://www.carnegiefoundation.org/classifications/index.asp?key=783)

Our Research Contributed To:

- AOIR Guidelines
- SACHRP Guidelines
- Numerous IRB Guidelines
- CITI Modules

Let’s Consider an Example:

- Investigator acquires/buys a data set from: (FB, Twitter, Geofeedia, etc)
  - Geofeedia is an aggregator and constantly mines “public” social media and Internet sites
  - “Location-based Intelligence”
- From the protocol, we know:

Data used for this analysis will come from multiple social media sources and delivered through the social media platform Geofeedia. Social media data from Twitter, Instagram, Facebook, YouTube, Flickr, Picasa, Sina Weibo, and VK will be mined for opioid- and heroin-related themes. Geofeedia provides a comprehensive collection of social media data that allows a user to search by location for a set of defined keywords, hashtags, and emoji's, etc. It also provides the ability to look backwards through time, as well as identify social networks by looking at how users interact and influence others in the social network. All posts are location-explicit and allow for easy mapping. Although all of the data used are publicly available, we will have the data deidentified (usernames anonymized and replaced with digital alpha/numeric code) and the locations will be be slightly offset by rounding the latitude/longitude coordinates by two significant digits.

Once the recordings have been initiated, we will geocode the crime data for the city of REDACTED. This will convert address-level data to latitude/longitude coordinates to be used for the spatial modeling. Geocoding the addresses and spatial modeling will be performed using ArcGIS 10.3. Local spatial clustering statistics such as Local Moran's I and Getis-Ord Gi*, will be used to detect clustering within the defined study sites. Once this is complete, social media-identified clusters will be mapped onto clusters of opioid and heroin-related arrests. These crime-related clusters will serve as a baseline and will be used for comparison with those clusters identified through social media.

The goal is to examine how well social media-derived clusters may serve as a proxy for real time surveillance of the epidemic. Here, we want to determine how well the social media-identified clusters overlap with those found in the crime data. One limitation may be that additional clusters may be discovered from those derived from social media and that these may not necessarily represent new, previously unknown clusters….

What Are the Major Issues?

IRB Considerations And Practices

- Ask investigators to include screen shots with protocols—can help in understanding data flows, recruitment processes, etc
- Consider: Data in use, at rest, in transit, and in deletion: Different ethical considerations and security measures; describe procedures (including safeguards for collecting, storing, processing subject data and data destruction) for minimizing potential risks to subject's confidentiality
- Learn the nuances between and among data management practices, including de-and re-identification; anonymized, coded, aggregated
    - Data sharing and data use agreements (NIH, NSF mandates); important for researchers to work with REBs/IRBs in planning for data sharing—raw data? Themes?
- Specify where and under what conditions individuals will have access to the data, what will be available and to whom (Air gap, clean rooms, data access levels))

If aggregated anonymized data will be made publicly available, consider whether subjects could be (re)identified, and what level of risk applies

- Reconsider minimal risk and everyday life:

-*“We anticipate that your participation in this study presents no greater risk than everyday use of the internet”*

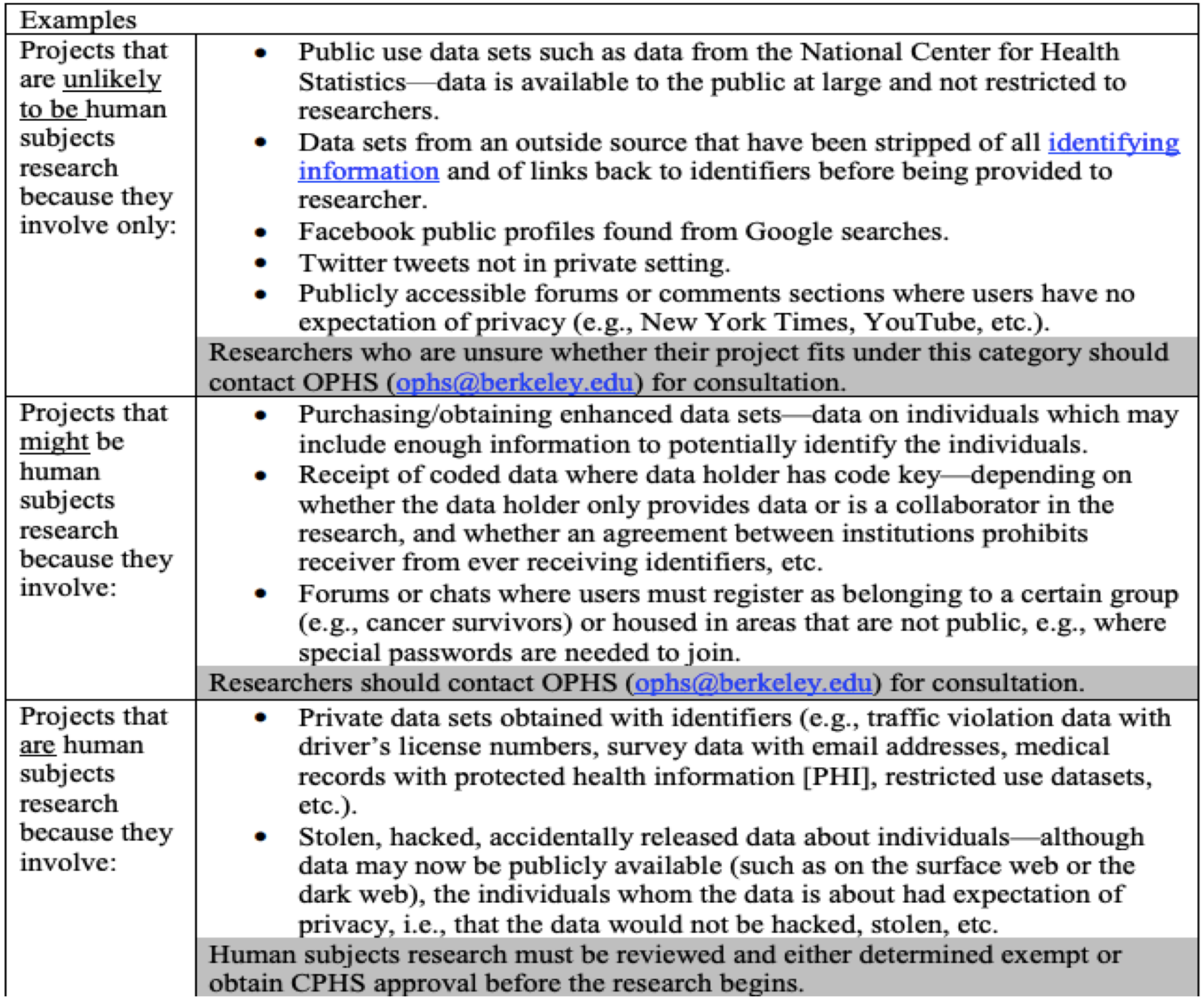
-*“Although every reasonable effort has been taken, confidentiality during actual Internet communication procedures cannot be guaranteed.”*

-*“Your confidentiality will be kept to the degree permitted by the technology being used. No guarantees can be made regarding the interception of data sent via the Internet by any third parties.”*

- Address uncertainty in data longevity in more open-ended terms: “Data may exist on back ups or server logs beyond the timeframe of this research project”
- Clarify that one’s consent to use, eg, Facebook, is not the same as consent to participate in research
- Ensure research is not in violation of TOS, user standards, norms
- Disclose what third party sites may be used for collection, storage, dissemination and that access by third parties is possible

- More IRBS using the following principles around identifiability and secondary data: Eg, UC Berkeley (https://cphs.berkeley.edu/secondarydata.pdf)

  - Research will NOT involve merging any of the data sets in such a way that individuals might be identified
  - Researcher will NOT enhance the public data set with identifiable, or potentially identifiable data
- The following uses of the data sets (such as …) may require prior IRB review or a determination of exempt status:
    - Merging data sets in such a way that individuals may be identified.
    - Enhancing a data set with identifiable or potentially identifiable data.
    - Research that consists of using one or more data sets on this list and also (1) the collection or use of private, identifiable data and/or (2) interactions or interventions with humans.



Examples	
Projects that are unlikely to be human subjects research because they involve only:	• Public use data sets such as data from the National Center for Health Statistics—data is available to the public at large and not restricted to researchers. • Data sets from an outside source that have been stripped of all identifying information and of links back to identifiers before being provided to researcher. • Facebook public profiles found from Google searches. • Twitter tweets not in private setting. • Publicly accessible forums or comments sections where users have no expectation of privacy (e.g., New York Times, YouTube, etc.).
	Researchers who are unsure whether their project fits under this category should contact OPHS (ophs@berkeley.edu) for consultation.
Projects that might be human subjects research because they involve:	• Purchasing/obtaining enhanced data sets—data on individuals which may include enough information to potentially identify the individuals. • Receipt of coded data where data holder has code key—depending on whether the data holder only provides data or is a collaborator in the research, and whether an agreement between institutions prohibits receiver from ever receiving identifiers, etc. • Forums or chats where users must register as belonging to a certain group (e.g., cancer survivors) or housed in areas that are not public, e.g., where special passwords are needed to join.
	Researchers should contact OPHS (ophs@berkeley.edu) for consultation.
Projects that are human subjects research because they involve:	• Private data sets obtained with identifiers (e.g., traffic violation data with driver's license numbers, survey data with email addresses, medical records with protected health information [PHI], restricted use datasets, etc.). • Stolen, hacked, accidentally released data about individuals—although data may now be publicly available (such as on the surface web or the dark web), the individuals whom the data is about had expectation of privacy, i.e., that the data would not be hacked, stolen, etc.
	Human subjects research must be reviewed and either determined exempt or obtain CPHS approval before the research begins.