Emerging Issues in Internet Research

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

Today’s Goals

• Discuss the (rapidly) shifting infrastructure and modalities of “the Internet” and the impact on research
• Consider the ways in which “personal” identities are no longer individuated, but shared (with consent or not)
• Review current research regulatory models and the challenges they are facing vis-à-vis “the new normal”
• Explore emerging challenges and opportunities around recruitment and consent

Internet Research Defined (SACHRP 2013)

• Research studying information that is already available on or via the Internet without direct interaction with human subjects (harvesting, mining, profiling, scraping)—observation or recording of otherwise-existing data sets, chat room interactions, blogs, social media postings, etc.)
• Research that uses the Internet as a vehicle for recruiting or interacting, directly or indirectly, with subjects (Self-testing websites, survey tools, Amazon Mechanical Turk®, etc.)
• Research about the Internet itself and its effects (use patterns or effects of social media, search engines, email, etc.; evolution of privacy issues; information contagion; etc.)
• Research about Internet users—what they do, and how the Internet affects individuals and their behaviors
• Research that utilizes the Internet as an interventional tool, for example, interventions that influence subjects’ behavior
• Others (emerging and cross-platform types of research and methods, including m-research (mobile))
• Recruitment in or through Internet locales or tools, for example social media, push technologies
What was the Old Normal?

The Good Old Days

- We went to “it”
- Oftentimes, we downloaded “it” (and maybe “owned” it) to our own device (we controlled what we got)
- Sometimes, we even downloaded it WITHOUT SIGNING A TOS/EULA! Does anyone remember anonymous FTP?
- We decided what we wanted to see, explore, do
- We could be “invisible” if desired
- We had to “connect” – connectivity was not ubiquitous
  – [http://www.youtube.com/watch?v=gsNaR6FrU0Q](http://www.youtube.com/watch?v=gsNaR6FrU0Q)

I MISS
MY
PRE-INTERNET
BRAIN
There is zero privacy anyway, get over it

Although you can upload your data using a pseudonym, there is no way to anonymously submit data. Statistically speaking, it is really unlikely that your medical and genetic information matches that of someone else. By uploading you do not only disclose information about yourself, but also about your next kinship (parents and siblings) that shares half of a genome with you. Before uploading any genetical data you should make sure that those people approve of you doing so. (openSNP)
Why? And When did this Happen?

Source: Fred Cavazza, Paris, France, 2009
"Intervention" as defined by DHHS regulations means both physical procedures by which data are gathered (for example, venipuncture) and manipulations of the subject or the subject's environment that are performed for research purposes. [45 CFR 46.102(f)]

"Interaction" as defined by DHHS regulations means communication or interpersonal contact between investigator and subject. [45 CFR 46.102(f)]

"Private information" as defined by DHHS regulations means information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record). [45 CFR 46.102(f)]

"Identifiable information" as defined by DHHS means information that is individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information).
Core and Derived Data

Some of My Core Data

What Does all of this Mean for Research—and Research Ethics?

• Blurring of boundaries or venues
• Researchers are but one stakeholder in a larger framework
• Interoperability
• Remote research—separation of researcher-subject
• Dissolving public v private sector data flows
• Non-consensual access to data
  — Pros and cons: DNA databases to identify family members in War Zones; potential of identifying families of war criminals?
  — Secondary subjects and third parties could benefit from rights to have their family members identified or could become more vulnerable and targeted
• Fundamental principles of autonomy, consent, privacy do not align
• Individuals are known by association or through a composite of data, true or untrue: your data persona
• How to isolate a research frame
• More emphasis on DATA SECURITY versus individual privacy
Extant Regulatory Models

- In the US: Do not address the confounding issues around HUMAN SUBJECTS research vis-à-vis today’s technological environment
  - Attempts (Buchanan and Gallant/SACHRP, 2013)
- Some guidelines in Canada
- EU focus on data privacy and ownership
- But research is now multi-national, trans-border, globally conducted. Who’s or what ethical models should apply?
- Context-dependent approaches seem most appropriate

Finding a Balance between the Regs and the Extra-Regulatory: Considerations and Language

- Consider: Data in use, at rest, and in transit: Different ethical considerations and security measures
- Explain how data are maintained. In individually identifiable form, aggregate form, anonymized?
  - Becoming more and more significant with data sharing and data use agreements (NIH, NSF mandates); important for researchers to work with IRBs in planning for data sharing!
- Don’t collect what you don’t need! (see next slide; some researchers are screaming at me!)
- Develop and use data security plan
  - See, for example Harvard, North Carolina, Oregon
- If aggregated anonymized data will be made publicly available, consider whether subjects could be (re)identified.

Best Practices

- Ask investigators to include screen shots with protocols—can help in understanding data flows, recruitment processes, etc
- Explain how data are transmitted: Is a survey host used? Will the host retain identifiable data? Will the data be encrypted? Is IP and other identifying information withheld from the researcher: (This is the “readily ascertainable” piece)
Best Practices

• Reconsidering minimal risk and everyday life: “We anticipate that your participation in this study presents no greater risk than everyday use of the internet” (Cornell)

• If aggregated anonymized data will be made publicly available, consider whether subjects could be (re)identified.

• Describe procedures (including safeguards for collecting, storing, processing subject data and data destruction) for minimizing potential risks to subject’s confidentiality:
  – AGAIN: “Although every reasonable effort has been taken, confidentiality during actual Internet communication procedures cannot be guaranteed.”

• Specify where and under what conditions individuals will have access to the data, what will be available and to whom. (Univ of Denver)

• Researchers should not collect “opportunistic data” (Cornell)

• Turk: Be sure to mention RESEARCH versus job: “This is an academic not-for-profit research study...” (Cornell)

Best Practices

• Identify potential tracking measures if using work or public computers or recommend use of private computers.

• Disclose what third party sites may be used for collection, storage, dissemination and that access by third parties is possible (we may not know that is ok!)

• “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.” (Penn State)

• Can apply to mresearch

• Need better standardized msecurity and encryption

• 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”

• Move away from “locked file in a locked office” mentality; the “cloud” or dispersed computing is the new normal!

Best Practices: Security Considerations

• Identify levels of data sensitivity (the sauerkraut protocol versus the eating disorder protocol)

• Develop risk categories and procedures

• Locked down versus open door data and access by investigators; what really needs robust protections? Air gap?

• PASSWORD PROTECT—and use robust passwords!

• Remote deletion and remote lock down capabilities

• Remote storage and backups

• Work with your IT and CIO! Collaboration is key.
Sample Language for Consent Documents

- “Although it is unlikely that anyone will try to gain access to your email, you have the right to know that email transmissions are not private and therefore transmission of information through this form cannot be guaranteed to remain confidential.” (Bard)
- “Please note that the online survey is hosted by Company ABC which is a web survey company located in the USA. All responses to the survey will be stored and accessed in the USA. This company is subject to U.S. Laws, in particular, to the US Patriot Act/Domestic Security Enhancement Act that allows authorities access to the records that your responses to the questions will be stored and accessed in the USA. The security and private policy for Company ABC can be viewed at http://…”
- (with mobile device research): “Remote data deletion will be performed in the event of a lost or stolen phone”

Sample Language for Consent Documents

- Glickman, McNair, et al, 2012: “If you participate in this clinical study, you should feel free to discuss the study with your family and with other people who are close to you. You should also tell any health care providers who treat you that you are in the study. However, to help make sure that the data from the study is as accurate and reliable as possible, please do not discuss information about the study in public places while the study is in progress. Public places may be situations like support groups, or may be places like internet message boards. If you have questions about side effects, please talk to your study nurse or study doctor.”

Selected Internet Research Guidelines

As researchers...

• “Classic consents must therefore transition away from attempting to guarantee individuals’ privacy. Rather, new forms of consent should aim at educating research subjects on what the data collected on them can say and the degree to which it can or cannot be protected” (Schadt, 2012).
As Researchers...

- Research is now an ENTERPRISE—complex and multifaceted. Team approach is best: computer scientists, security, ethicists, methodologists.
- Good ethics = good science and good research
- Let ethics drive methods

Contact Information

Elizabeth Buchanan, PhD
buchanane@uwstout.edu
715.232.2477
@stout_ethics
@stout_research