OHRP Webinar Lecture Series

Conducting Internet Research: Challenges and Strategies for IRBs

Assessing Privacy and Identifiability, and Maintaining Confidentiality

Laura Odwazny
Senior Attorney
Office of the General Counsel
U.S. Department of Health and Human Services
DISCLAIMER

This presentation does not constitute legal advice. The views expressed are the presenter’s own and do not bind the U.S. Department of Health and Human Services or its operational components, including the Office for Human Research Protections.
OUTLINE

- What is Internet research
- How specific requirements of the HHS protection of human subjects regulations apply to research using the Internet
- Discussion of challenges in managing the ethical issues and regulatory considerations, focusing on assessing privacy and identifiability of subject information, and maintaining confidentiality.
- Possible developments on the Federal horizon relevant to Internet research
SETTING THE STAGE: WHAT IS INTERNET RESEARCH?

- Internet research
  - Internet used as a tool for conducting research
    - Examples: online survey, subject recruitment, email or chat interviews
  - Internet as a location or site for conducting research
    - Examples: Collecting data about or observing online environments such as chatrooms, gaming sites, virtual worlds
  - Internet as a source of information
    - Examples: data mining from social media site; collecting data from online datasets, databases, repositories

“Recommendation Concerning Internet Research and Human Subjects Research”
What Types of Internet Research Do IRBs Encounter?

- As of 2007, IRBs reviewed:
  - Online Survey Research (98%)
  - Online Ethnography (1%)
  - Other (Data sets) (1%)


- Times have changed! What IRBs encounter now:
  - Data-scraping bots, mechanical turks, virtual dentistry education simulation, subject recruitment/retention via social media, online clinical trials
WITH INCREASING INTERNET SATURATION…

Social media – Internet-based applications that allow creation and exchange of user-generated content

Provide mechanisms for users to interact:
--chat, instant messaging, email, video, file sharing, blogging, discussion groups
PLUS THE GROWING AVAILABILITY OF BIG DATA...
…BIG DATA THAT MAY BE IDENTIFIABLE…

GENETIC DATA + AGE + REGION, COMBINED WITH GENEEOLOGY WEBSITE AND GOOGLE SEARCHES = 5 INDIVIDUALS (AND THEIR FAMILY MEMBERS) IDENTIFIED
= Increasing Use of Internet for Research

Note:

- The HHS protection of human subjects regulations do not specifically reference Internet research.
- OHRP has no formal written guidance specifically on Internet research.
March 13, 2013: SACHRP voted to provide recommendations re internet research

- SACHRP = Secretary’s Advisory Committee on Human Research Protections
- SAS and SOH subcommittees developed recommendations for SACHRP to make to Secretary of HHS and Assistant Secretary of Health re internet research
- Recommendations are not official OHRP guidance, as not yet adopted by HHS or OHRP
IN THE ABSENCE OF SPECIFIC INTERNET REGULATIONS/GUIDANCE…

Apply the existing regulations and OHRP guidance!

Question for contemplation: How different is Internet research from other types of research? Is it special?
Some of the Big Regulatory Issues Related to Internet Research

- What is “private”?
- What is “identifiable”?
- How to protect subjects’ privacy and confidentiality interests?
- Minimizing risk when using sensitive online data
  - Current sensitivity vs. future sensitivity
SOME OF THE RELATED REGULATORY DECISION POINTS

- Is the activity research?
- Does the research involve human subjects?
- Does the human subjects research qualify for exemption from the regulatory requirements?
- Does the research present no more than minimal risk such that it may be reviewed via expedited review (if it meets a category)?
- Informed consent – obtained or waived/adjusted? How to describe confidentiality protections?
WHAT IS RESEARCH?

- Research: systematic investigation designed to develop or contribute to generalizable knowledge (45 CFR 46.102)
- Studying Internet sites or using Internet as a research tool
  - Studying online social networks
  - Online context as ethnographic field site (chat rooms, gaming research)
  - Data mining/scraping from Internet sites
  - Web-based surveys
  - Web-based interviews
HUMAN SUBJECTS – IDENTIFIABLE PRIVATE INFORMATION

45 CFR 46.102(f): “a living individual about whom an investigator conducting research obtains (1) data through intervention or interaction with the individual, or (2) identifiable private information

- Private information: “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).”
PRIVACY ON THE INTERNET?

How to interpret “reasonably expect that no observation or recording is taking place” or “reasonably expect will not be made public”

- IM, tweet, email, Facebook profile, chatroom discussion, listserve posting – what is reasonable expectation of privacy in each?

- Or is everything on the Internet that I can see public?
WHEN IS AN EXPECTATION OF PRIVACY “REASONABLE”?  

- People in online environments that are presumptively public often act as if they are in private space  
  - Caused by online feelings of anonymity, norms of the Internet space, reduced inhibitions, separation of people from text  
- Expectations of privacy may not equate with reality of privacy (or lack thereof)

Asa Rosenberg, “Virtual world research ethics and the private-public distinction,” International Journal of Internet Research Ethics, v.3, December 2010:  
http://ijire.net/issue_3.1/3_rosenberg.pdf
How may the IRB assess whether information obtained via the Internet should be considered private?

- Regulatory standard of “reasonable” does not depend on individual subject’s own expectation of privacy
- How to consider what expectations of privacy in the information are “reasonable”
  - Get information about the environment
  - Get information about the users
  - Review Terms of Service, site policy
Identifiable

Individually identifiable = subject’s identity readily ascertainable by the investigator or associated with the information

Structure of social network, search terms, purchase habits, movie ratings on Netflix may uniquely identify individual

- Zip code + sex + DOB enough for Professor Latanya Sweeney to uniquely identify 87% of US population (de-identified medical data linked to voter info re-identified patients by name)

Question for contemplation: given demonstrated ability to reidentify individuals from anonymized or aggregated data, is this a meaningful decision point?
**HOW CAN THE IRB ASSESS IDENTIFIABILITY?**

- When will the subject’s identity be “readily” ascertainable by the investigator or associated with the information?
  - Consider the investigator, e.g. Professor Latanya Sweeney vs. Professor Laura Odwazny
  - Consider the potential identifiers or partial identifiers
    - Direct quotes easily traceable to Twitter account even if handle is removed
  - Consider likelihood of reidentification with triangulation, not just whether it is theoretically possible
Avatars

Is information obtained via an avatar information about a human subject?
-- Human/bot?
-- Interaction/intervention?
-- Private and identifiable?

Sensitivity of information obtained from avatar observation akin to information obtained by observing humans?
RELEVANT EXEMPTIONS – ONLINE EDUCATION

- 45 CFR 46.101(b)(1): Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

- Internet locale could be an “established or commonly accepted educational setting” and online education could be a “normal educational practice”

- Examples:
  - Evaluating the conduct of a web-based class
  - Assessing the efficacy of the use of social media site to disseminate class information
  - Comparison of virtual simulation training to traditional training – ex/ online dentistry procedures conducted in Second Life
RELEVANT EXEMPTIONS – EDUCATIONAL TESTS, SURVEY AND INTERVIEW RESEARCH, OBSERVATION OF PUBLIC BEHAVIOR

- 45 CFR 46.102(b)(2), unless: information is recorded in a manner whereby subjects can be identified **AND** disclosure of the responses could reasonably place them at risk of criminal or civil liability or be damaging to the subjects’ financial standing, employability, or reputation.

- What is “recorded in a manner whereby subjects can be identified” when the Internet is used?

- What is “observation of public behavior” online?
RELEVANT EXEMPTIONS – DATA MINING

- 45 CFR 46.101(b)(4) -- collection or study of existing data/specimens, if sources are publicly available or if information is recorded by investigator in such a manner that subjects cannot be identified
  - When is information “recorded in an identifiable manner”?
  - When are data, documents, or records publicly available on the Internet?
    - Does “publicly available” include large datasets purchased/obtained from Google or Facebook?
    - What if data are restricted -- available only to ‘friends’, listserv members?
EXEMPTION 4 CONTINUED: “RECORDED IN A MANNER WHEREBY SUBJECTS MAY BE IDENTIFIED…”

- Is an email address an identifier?
- Do tweets contain identifiers?
- Does the inclusion of IP address make information identifiable?

Note: For HIPAA, OCR has stated position (below); OHRP has no formal guidance

The second is the “Safe Harbor” method:

(2)(i) The following identifiers of the individual or of relatives, employers, or household members of the individual, are removed:

- (E) Fax numbers
- (F) Email addresses
- (G) Social security numbers
- (M) Device identifiers and serial numbers
- (N) Web Universal Resource Locators (URLs)
- (O) Internet Protocol (IP) addresses
If Not Exempt… IRB Review

Challenges in IRB review of Internet research:

- Requirement that risks be minimized
  - Two main sources of risk:
    - Participation -- No direct contact with subjects; more difficult to deal with individual reactions (intervention, debriefing, follow-up)
    - Breach of confidentiality

- Eligible for expedited review?
  - Must be minimal risk and fall within expeditable research category
MINIMAL RISK

- Probability and magnitude of harm/discomfort in the research not greater than ordinarily encountered in daily life or during routine physical or psychological examinations/tests (46.102(i))
  - Gateway to expedited review; waiver of consent and documentation; no need to explain compensation or any treatments for research-related injury in consent; Subparts B, C, D categories of permissible research

- Risks associated with data security breach, likelihood of access by 3rd parties alter conception of minimal risk in Internet research?
  - Less privacy, more observation in general in daily life
INTERNET-BASED SUBJECT RECRUITMENT

- Facebook page
- YouTube video
- Matching algorithm on social media sites (e.g., PatientsLikeMe)
- "Push" method (e.g., Inspire.com)
OHRP GUIDANCE ON SUBJECT RECRUITMENT

- OHRP considers subject recruitment part of informed consent
  - Recruitment plan must receive IRB review/approval prior to initiation


- No IRB review needed for descriptive information:
  - study title
  - purpose of the study
  - protocol summary
  - basic eligibility criteria
  - study site location(s)
  - how to contact the study site for further information
OHRP Guidance (continued)

- IRB review needed if additional information provided
  - Description of research risks/potential benefits
  - Solicitation of identifiable private information (e.g. eligibility survey)
  - Incentives – monetary and non-monetary

- What needs to be reviewed:
  - Recruitment plan, not the actual webpage
    - But screen shots may be helpful to the IRB
RECRUITMENT VIA YOUTUBE VIDEO

Triple-Negative Breast Cancer Clinical Trial

Triple-negative Breast Cancer Clinical Trial Calling for Participants
USING SOCIAL MEDIA FOR RECRUITMENT – MATCHING TOOL

Research Tools

Clinical Trials Tool
We've integrated PatientsLikeMe with ClinicalTrials.gov to develop a clinical trials matching tool. This allows you to find trials you might be eligible for (including trials of drugs, devices, therapy, or non-interventional studies such as genetics or questionnaires) based on your age, sex, conditions, and location. Click now to find out about trials for patients like you.

ALS Lithium Study - The Results
In 2008, a small Italian study was published suggesting that the drug Lithium could slow the progression of ALS. In response, hundreds of members of PatientsLikeMe began taking the drug and using a new tool and a matching algorithm to conduct a patient-led observational study. The results of this study, published in Nature Biotechnology, show that we were unable to replicate the promising findings of the Italian group, but that PatientsLikeMe may provide a useful way of conducting observational studies faster and cheaper than existing trial methods.

PatientLikeMe Research
In collaborations with patients, academics and industry leaders, the research team designs and runs studies that contribute to the understanding of each of the conditions we serve and the evolution of the site itself. This new research tab will be where we will collect and present our patient tools and research findings. If you would like to contribute ideas about the design of the site or this page please let us know through the suggestions box or link to send the research team an email.

Meet the Research Team
The research team brings a variety of expertise to the design of PatientsLikeMe and the study of the data collected here. Our team's specialties include genetics, sociology, psychology, nursing, drug discovery, predictive modeling, and user behavior.
SOCIAL MEDIA AS RECRUITMENT TOOL: “PUSH” METHOD

Blog post from the founder of Inspire.com (3/1/09):

“I'm writing today to let you know about some new features we're introducing related to clinical trials... What's new is that from time to time we'll tell you about clinical trials in which you may be interested in participating. If you're not interested in participating, simply do nothing. If you do think you might be interested, we'll provide a link where you'll be able to read about a trial, decide if you are interested in participating, and fill out a short survey to see if you may qualify. If it appears that you may qualify, we'll put you in touch with the physicians conducting the trial so that you can learn more and find out if you do qualify.”
CONSIDERATIONS WITH USE OF SOCIAL MEDIA FOR RECRUITMENT

- Nature of social media data – easily transmitted quickly within and outside of social network
  - If recruitment method can identify an individual, any potential downstream harms?
  - What happens if recruitment information goes viral?
- Uncontrolled following discussion among viewers/bloggers: interactive, not static
  - Subsequent posts in effect add to posted information from user perspective?
- Must PI/IRB actively monitor social media sites used for recruitment for accuracy of information posted in comments, information about possible unanticipated problems?
INFORMED CONSENT IN INTERNET RESEARCH
CONSENT PROCEDURES

- Consider waiver of consent and/or documentation, if appropriate
- A “portal” can be used to provide consent information.
  - Subjects must click through consent page to get to survey
- Where documentation required – consider alternatives to traditional documentation
  - Electronic signatures (state and local law dictate acceptable form)
  - OHRP FAQ on electronic signatures: http://answers.hhs.gov/ohrp/questions/7260
Enroll in 2013 Study

Welcome to the OWIUS website!

You should only be reading this if you are a student at Otago University this year and/or in 2012 and have followed the link here.

To enroll in the study, please read the Information Sheet and Consent Form which follow this page. On submitting the consent form you will receive an email with your own user ID# and the link for the online well-being survey for this study.

Enrolment and access to the online survey will close on 30 September 2013.

Thank you.
CHALLENGE: PROTECTING CONFIDENTIALITY WHILE OBTAINING CONSENT IN INTERNET RESEARCH

- Sometimes no direct researcher – subject interaction
  - Interaction could be through avatar, profile, survey tool
- Not always clear who subjects are
  - Fluidity of group membership, identity assumed online may differ from actual identity
- May not be desirable or feasible to obtain documentation of consent
  - May provide more identifiable subject information than necessary (could increase risk); fluid group membership, e.g. chat rooms
  - Subjects may be surveilled unknowingly to them or the researcher (key stroke monitoring, spyware)
  - Digital maleficence
DESCRIPTION OF CONFIDENTIALITY PROTECTIONS IN INFORMED CONSENT

- 45 CFR 46.116(a)(1)(5) – informed consent must include statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained.

- “Locked file cabinet in locked room” description not sufficient for Internet research!

- Regulatory requirement pertains to “identifying” records: consider potential identifiability of research data obtained using the Internet.
**Consider When Describing Confidentiality Protections Including...**

- **How subject information is transmitted via the Internet**
  - Survey host (e.g., Zoomerang, Survey Monkey) used? Will host retain identifiable information? Will the transmission be encrypted?

- **How information is maintained**
  - Individually identifiable form, de-identified aggregate form?
  - Cloud storage?

- **Circumstances in which subject information might be disclosed outside the research team**
  - Data sharing and data use agreements increasingly being required by funding agencies (NIH, NSF mandates)
  - Remember funding agency access rights and possible mandatory disclosure to OHRP, FDA, ORI, other oversight agency
  - Patriot Act allows access to cloud
CONSIDER WHEN DESCRIBING CONFIDENTIALITY PROTECTIONS (2)

- **Data security plan**
  - Explain the efforts to protect the data, e.g., secure servers, computers not connected to university network

- **Do not absolutely guarantee confidentiality of subject information**
  - Unrealistic and likely inaccurate

- **If aggregated de-identified data will be made publicly available, consider the likelihood of re-identification of individual subjects whether this should be described**
On the horizon...
ANPRM seeking comment on possible areas of change to the Common Rule

Published July 26, 2011 by HHS “in coordination with the Office of Science and Technology Policy”: “Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay and Ambiguity for Investigators”

1000+ comments received
FEDERAL RULEMAKING PROCESS

HTTP://WWW.REGININFO.GOV/PUBLIC/REGININFO/REGMAP/INDEX.JSP

The Reg Map
Informal Rulemaking

Step One: Initiating Events
- Agency initiatives
- Legislative initiatives
- OMB initiatives

Step Two: Determination Whether a Rule Is Needed
- Agency initiatives
- Legislative initiatives
- OMB initiatives

Step Three: Preparation of Proposed Rule
- Administrative Procedure Act
- OMB Review of Proposed Rule
- OMB Review Under Executive Order 12866
- Final Rule

Step Four: OMB Review of Proposed Rule
- Administrative Procedure Act
- Final Rule

Step Five: Public Comments
- Administrative Procedure Act
- Final Rule

Step Six: Preparation of Final Rule, Interim Final Rule, or Direct Final Rule
- OMB Review of Final Rule
- Interim Final Rule
- Direct Final Rule

Step Seven: OMB Review of Final Rule
- Final Rule

Step Eight: Publication of Final Rule
- Final Rule

Optional Supplementary Procedures to Help Prepare a Proposed Rule
- Advance Notice of Proposed Rulemaking
  - An advance notice of proposed rulemaking requests information needed for developing a proposed rule.

↑ OHRP is in Step 3
ANPRM-- IMPLICATIONS FOR INTERNET RESEARCH

- To protect from informational risks (inappropriate use/disclosure of information), mandatory data security measures “modeled on” HIPAA?
- Apply Common Rule to all institutions receiving support from CR agency?
- No continuing review for most minimal risk research?
ANPRM – PROPOSALS FOR “EXCUSED” RESEARCH

- Add a new category of minimal risk SBR involving competent adults?
- Additional requirements for “excused” (formerly exempt) research?
  - Consent, oral or written, depending, with waiver contemplated
    - Oral w/o documentation for educational tests, surveys, focus groups, interviews
  - Data security standards
**Timeframe for NPRM?** As of April 10, 2014, Fall 2013 Regulatory Plan includes...

<table>
<thead>
<tr>
<th>Action</th>
<th>Date</th>
<th>FR Cite</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANFRM</td>
<td>07/26/2011</td>
<td>78 FR 44512</td>
</tr>
<tr>
<td>ANFRM Comment Period End</td>
<td>18/26/2011</td>
<td></td>
</tr>
<tr>
<td>NPRM</td>
<td>04/04/2014</td>
<td></td>
</tr>
</tbody>
</table>

Additional Information: Includes Retrospective Review under E.O. 13563.

Regulatory Flexibility Analysis Required: Undetermined

Government Levels Affected: Undetermined
THANK YOU AND STAY TUNED!

QUESTIONS FOR OHRP?
TOLL-FREE : (866) 447-4777
E-MAIL: OHRP@HHS.GOV