Considerations and Recommendations Concerning Internet Research and Human Subjects Research Regulations, with Revisions

Final document, approved at SACHRP meeting March 12-13, 2013

Introduction

The purpose of this document is to provide a starting point for the development of FAQs and/or Points to Consider regarding the conduct and review of Internet research. Current human subjects regulations, originally written over thirty years ago, do not address many issues raised by the unique characteristics of Internet research. Some IRBs, concerned about their ability to make appropriate and responsible decisions regarding Internet research, have developed working guidelines for investigators.[1] Many of these guidelines focus on technical questions about data security, but there are other issues to address: basic categorizations of types of Internet research; types of data; data identifiability and subject privacy; appropriate consent and authentication of subjects procedures; jurisdictional authority; data collection, data storage, research administration, and data destruction; data sharing practices and implications; and discussion of what is common, reasonable, and acceptable in a given Internet environment and how these standards relate to current regulations and guidance in the areas of informed consent, recruitment, and risk of harm.

Ethical conduct of Internet research also brings questions of scientific design into high relief: authenticity of subject identity, assurance of comprehension of consent, and verification of data integrity can present significant challenges.

Forms and Examples of Internet Research

There are multiple forms of Internet research. Some experiments are conducted fully in online fora or conditions; some research may include elements conducted through the Internet, for example, using a social media application as a recruitment tool combined with traditional research methods and spaces; some research can only be conducted on the Internet, for example, an ethnography of an online-only forum that has no corresponding geo-physical location; or, the Internet may be a tool underlying data collection. We identify a range of Internet research where human subjects may be involved:

- Research studying information that is already available on or via the Internet without direct interaction with human subjects (harvesting, mining, profiling, scraping[2]—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

[1] See for example: http://irb.uconn.edu/Internet_research.html
—http://www.marianuniversity.edu/interior.aspx?id=13714
—http://inside.bard.edu/irb/guidelines/
—http://www.luc.edu/irb/irbonlinesurveys2.shtml
—http://www.research.psu.edu/policies/research-protections/irb/irb-guideline-10

[2] Terms that may be unfamiliar are highlighted in blue bold and included in the Glossary.

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))[3]

- Recruitment in or through Internet locales or tools, for example social media, push technologies

The broad and overarching term "Internet research" includes both the Internet as a *tool for research* and the Internet as a *locale or venue of research*. For example, research employing survey instruments, search engines, databases, databanks, or aggregators would constitute using the Internet as a tool for research. Such research may not involve direct interaction with human subjects, but identifiers or personally identifiable information may be generated, collected, and/or analyzed. In contrast, using the Internet as a medium or locale of research entails qualitative or quantitative studies of various Internet "spaces," such as chat rooms, gaming worlds, virtual environments, or other simulated locales. Internet phoning, video conferencing, or online chat may be both tool and venue; applications such as Skype® or Facetime® may be used to contact subjects or participants, and interviews or focus groups can be conducted via the application. The increasing predominance of social media, defined as a "group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content,"[4] is blurring the tool-versus-venue model. Consider, for example, research using a social media application that engages subject recruitment via targeted ads on such platforms as Facebook® or via microblogging tools such as Twitter®, uses online data collection and analytic tools, and disseminates data via other social media applications. A specific example comes from an ongoing exploratory group of the ASCO Integrated Media and Technology Committee, which is reviewing the regulatory, legal, and ethical implications of oncology research and social media usage.[5] Projects using community-based participatory research methods are embracing Internet and m-research to send, receive, collect, and disseminate data synchronously. Other examples include such applications as CenceMe®, which integrates with social media and cellular devices to "infer a person's activity…and social context. This information is shared within the person's social circle…."[6]

[3] Much discussion is occurring around FDA approval of mobile applications for medical research**.** In July 2011, the FDA released draft guidance for Industry and Food and Drug Administration Staff - Mobile Medical Applications available at http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm263280.htm#5

[4]Kaplan, Andreas M.; Michael Haenlein (2010) "Users of the world, unite! The challenges and opportunities of Social Media". Business Horizons 53(1): 59–68.

[5]Dizon, D. *et al*. (2012). Practical guidance: The Use of Social Media in Oncology Practice. *Journal of Oncology Practice, 000610.*

[6]Miluzzo, E. *et al.* (2010). Research in the App Store Era: Experiences from the CenceMe App Deployment on the iphone." *ACM, 978-1-60558-843-8-10/09*

Clear boundaries between"grid-enabled"technologies are eroding. For example, mobile applications interface with Internet sites or venues; tablets connect with cloud-based services in the use of survey tools; mobile devices are used in conjunction with Internet-enabled methods such as momentary sampling, reverse RSS data feeds, or synchronous data collection and analysis. With the emergence of such cross-operational research, fundamental aspects of human subjects research (recruitment, informed consent, data identifiability) present new challenges. Consent, for example, may occur in a synchronous setting, where both investigator and subject share a virtual space; or, consent may occur asynchronously, where a consent form is posted and subjects review it in the absence of the investigator. In the latter case, best practices are needed to ensure appropriate comprehension of consent documents and processes.[7]Thoughtful IRB review of emerging forms of cross-platform, cross-operational research may increasingly demand technical expertise in addition to regulatory knowledge, as new methodologies complicate risk/benefit analyses, and raise issues of confidentiality, privacy, and voluntariness.

This document argues for a reasoned and balanced approach to review of Internet research protocols, and does not advocate for more stringent review of Internet research. Nevertheless, in some circumstances researchers/investigators may have additional responsibilities. The ease with which sensitive data can be accessed, shared, hacked, and/or replicated is unique to Internet research, and for this reason, investigator responsibilities for good data stewardship, and heightened awareness of subjects' privacy, confidentiality, and identities, are critical.

Recommendations

This document is based on input from the research and professional literature, multiple years of workshops at Public Responsibility in Medicine and Research (PRIM&R) Advancing Ethical Research conferences, on-site panels at SACHRP in 2010, 2012, and 2013, and meetings with members of the SAS and SOH subcommittees of SACHRP in 2012 and 2013.Based on these experiences, we recommend that OHRP , the Food and Drug Administration (FDA), and the Office for Civil Rights (OCR) jointly commit to producing formal FAQs or Points to Consider for the research and IRB community that address the issues presented below. In addition we recommend that IRBs be provided with lists of appropriate questions to ask when reviewing Internet research, lists of appropriate or acceptable characteristics of vetted third party tools,[8] terms and phrases to use in protocols and informed consent/information sheet documents, a glossary of terms frequently used in Internet research,[9] as well as a decision-making flow chart that resembles existing models. Most of these are still in development.

Fundamental Principles

Investigators and IRBs should remember that the Belmont Report's fundamental principles of respect for persons, beneficence, and justice are as applicable to Internet research as they are to any other form

[7] For example, electronic comprehension checks (quizzes or short responses) of a consent document may be embedded prior to a subject's entrance into a study site or prior to engaging in any research activities. In some studies, subjects must score 100% accuracy on their comprehension checks to be eligible for the research.

[8] Given the frequency with which commercial tools change their terms of service, this list would of necessity be based on appropriate characteristics, rather than calling out specific companies or products.

[9] See Appendix A (FORTHCOMING)

of human subjects research. Regardless of how the regulations may be interpreted in individual studies, adherence to these fundamental principles is important to encouraging public trust in the ethical conduct of Internet research.

Regulatory Considerations

Q1: What is "research involving human subjects" on the Internet?

The regulatory definitions of *research*, *human subject,* and *identifiable private information* must be our starting points.

> *Research* means a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge.
>
> *Human subject* means a living individual about whom an investigator (whether professional or student) conducting research obtains
>
> (1) Data through intervention or interaction with the individual, or
> (2) Identifiable private information.
> …
> *Private information* includes 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). Private information must be individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects.[10]

Use of the Internet as a tool for research and for intervention or interaction with subjects does not, in general, challenge the definition of "human subject." However, new forms of identity, including avatars or other Internet personae, can be considered as virtual representations of "human subjects" if personally identifiable information about living individuals can be obtained by observing the actions of, or interacting with, the avatars. Investigators should determine if the avatar is a proxy for the individual, and if so, whether the subject's personally identifiable information (PII) is being obtained. Some avatars, for example, are simply computer-generated characters or representations and have no connection to an individual's PII. Other forms of Internet personae, including *bots*, typically do not display PII. Bots may be programmed to mine or harvest discrete PII from individual profiles, web sites, etc., or they may harvest large collections of data, such as patterns of search behaviors. If a bot's purpose is to collect and display an individual's PII, it may itself be a proxy for a "human subject." If its purpose is to harvest multiple individuals' PII from multiple sources, its activity might constitute human subjects research (but the bot itself would not be considered a human subject). Depending on the nature of the data and how they are obtained, these entities' activities may or may not require IRB review. (See also footnote 11)

The issues of "identifiability" and of "private information" are addressed below (see Q3 and Q4).

[10] 45 CFR 46.102(d) and (f)

Q2: What is exempt research involving human subjects on the Internet?

The use of the Internet to deliver educational materials is now common, and the regulatory exemption at 46.101(b)(1) for certain types of education research will often apply. (See Q8 for further discussion of "normal educational practice.") The exemption at 46.101(b)(2) for certain kinds of tests, surveys, interviews, or observation of public behavior, where the collected information is not sensitive or is not identifiable, is much more complicated; the complications hinge on the issues of "public behavior" and "identifiable." (See Q3, Q4, Q6, and Q7.) The exemption at 46.101(b)(4) raises similar questions about "publicly available" information and the identifiability of subjects, which are addressed below at Q3 and Q4.

Q3. When, if ever, is information available via the Internet "private information," and when can subjects "reasonably expect" their private information will not be made public?

Private information as defined in the Common Rule 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)]

If individuals intentionally post or otherwise provide information on the Internet, such information should be considered public unless existing law and the privacy policies and/or terms of service of the entity/entities receiving or hosting the information indicate that the information should be considered "private." To the extent that terms of service or explicit prohibitions would preclude the use of data on the Internet for research purposes, the determination that such data should be considered "private" is clear. In addition, investigators should note expressed norms or requests in a virtual space, which – although not technically binding – still ought to be taken into consideration.[11] When in doubt about whether to consider data public or private, investigators are encouraged to consult with their IRB about the specific circumstances. IRBs should be aware of changing terms, site security, and information/data use policy. For example, what was once considered private information may change based on the business model of the site.

(a) Are there now, or should there be, consensus standards for privacy of information on the Internet? The regulatory definition of private information cites medical records as an example. Tax records

[11] For example, "Everyone is welcome on PrettyThin. Anorexics, ex-anorexics, people in the health profession….it's an open forum. The alternative is the closet…is that the society we wish to live in?" (http://www.prettythin.com/category/frequently-asked-questions/). However, a different approach is offered at Ana Boot Camp: "Some images, links text and thinspiration may be considered triggering in nature. As well, if you are looking to get anorexia / bulimia by being here then **please leave now**. You will not find information contained within this web site, forum, or any site linked to / from this website on how to become anorexic or bulimic.
If you do not accept the condition of anorexia / bulimia / other eating disorders plus the pro-ana pro-mia movement then you must also leave this proana website immediately. Also you will not use this pro-ana pro-mia web site and or forum against anyone in any conceivable manner. You have been forewarned. By entering this proanapromia web site you are signing a digital certificate stating that you have read and understand the above mentioned conditions and you are entering this proanapromia site knowingly and willingly of the aforementioned conditions. Entering by any other circumstance is perjury and can be punishable by law." (http://anabootcamp.weebly.com/)

or personal diaries are often also given as examples. Is it possible to define categories of information on the Internet that are, by default, private and others that are, by default, public? For instance, at one extreme, identifiable information that is available only with a subject's permission, or by using a password or other access mechanism under the subject's control, could be considered private. At the other extreme, information that is legally available to any Internet user, without special authorization or access permission, could be considered public.

(b) A subject's own expectation of privacy is not always "reasonable." A subject may assume—perhaps in ignorance—that his or her information provided or available on the Internet is private, but the first part of the regulatory definition of "private information" specifies that the individual "can reasonably [sic] expect that no observation or recording is taking place." Information that is archived online has, *ipso facto*, been recorded. Can it ever be *reasonable* to expect otherwise, absent an explicit statement that no information will be recorded?

(c) Despite (b) above, the Belmont principle of beneficence may support a more conservative approach. A subject who incorrectly assumed his/her identifiable information was private, or restricted only to a select group, might not have posted the information on some social networking site if s/he thought the information would be widely available, believing that the information could be embarrassing or damaging. Should the investigator and the IRB consider the proposed research to be subject to IRB review, even if under existing regulations the research is exempt because the information is publicly available? Researchers and IRBs should consider the nature of the study and the sensitivity of identifiable data; more details about the study, and thoughtful institutional policy, taken in consideration with standard professional or disciplinary norms and practices, would help inform such decisions.

(d) The second part of the definition cites a reasonable expectation that information provided for a specific purpose will not be made public. When is an online venue, or social or professional networking site, or other online activity considered "public"? Does it matter if a password is required to join the venue? If the venue is moderated? If the venue is intended for use by individuals who share a particular condition or interest? Are there "shared priorities" by the members that dictate or determine norms?

One suggestion would be to follow the published privacy/confidentiality policy of the site; if there is no policy the site could be considered public. Privacy policies may parallel "anonymous" meeting standards (e.g., Alcoholics or Narcotics or Gamblers Anonymous), where members operate according to a set of shared priorities and there is an expectation of privacy and confidentiality within *and* outside of the meeting. Investigators should be aware of and respect those shared expectations.

A less nuanced approach would be to say that any venue where membership or participation must be authorized should be considered private. In contrast, venues where any individual can participate without third party approval—even if a password (of the individual's own choosing) is required—would not be considered private. In addition, sites whose purpose is to present participants' comments for public review (such as the comment section follow a news article) would be considered public even if participants must be vetted or authorized to participate.

In addition to the above considerations under the Common Rule, a researcher may need to consider whether the entity receiving or hosting the individual's information is subject to the HIPAA Rules, and whether the information being maintained is protected health information. For example, a HIPAA

covered entity may offer a web-based personal health record to individuals. The information entered into the record by either the covered entity or individual would be protected health information under the HIPAA Rules and thus, may only be accessed for research or other purposes as permitted by the Rules.

Note that OCR would not consider research solely on information from publicly available sources (even if some health information was included) to be research involving protected health information for HIPAA purposes. However, a HIPAA covered entity that collects information about an individual from the internet and integrates the information into an individual's medical or other health record must now protect that information as it would the rest of the record.

Q4: What is *identifiable private information* on the Internet?

What are ways that identifiable private information may be conceptualized on the Internet?

Private information is considered *identifiable* under the Common Rule if "the identity of the subject is or may readily be ascertained by the investigator or associated with the information."[45 CFR 46.102(f)].The nature of online data enables mining and matching, raising the potential for partial identifiers to be combined and individuals recognized.[12] Multiple data sets may be aggregated and analyzed, yielding surprising or novel information. (Existing guidance and best practices regarding genetic databases and biospecimen banks may help inform thinking on these issues.[13])

If the identity of the subject cannot be "readily ascertained by the investigator or associated with the information" then the activity is not research involving human subjects. Unfortunately, the phrase "readily ascertained" is not defined, either in regulations or in guidance, and it is unclear whether the modifier "readily" also applies to "associated with the information." Trying to quantify "readily" in some way, such as "number of clicks needed to get to a name/identity," seems unlikely to be satisfactory.[14] Ultimately, IRBs must make the difficult determination as to when a researcher who obtains multiple data points about a person that, although not individually sufficing to identify the person, may in aggregate render the data readily identifiable.

[12] See for example, Sweeny, L. (forthcoming, *Connecting Your Dots: What they know from what you leave behind*; 2004; Privacy-Enhanced Linking. *ACM SIGKDD Explorations* 7(2) December 2005.
Earlier version available as Carnegie Mellon University, School of Computer Science Technical Report CMU-ISRI-05-136. Pittsburgh: November 2000); Narayanan, A. and Shmatikov, V. (2008). Robust De-anonymization of Large Sparse Datasets. In Proc. of *29th IEEE Symposium on Security and Privacy*, Oakland, CA, May 2008, pp. 111-125. IEEE Computer Society.

[13] See, for example, Report of the Public Responsibility in Medicine and Research (PRIM&R) Human Tissue/Specimen Banking Working Group Part I Assessment and Recommendations, 2007
http://oba.od.nih.gov/policy/Tissue%20Banking%20White%20Paper%203-7-07%20final%20combined.pdf

[14] If an IRB considers the expertise or qualifications of the investigator when considering how "readily" an identity can be ascertained, then depending upon the investigator's skill and experience, and availability of other data, use of the same information by two different investigators could in one case constitute research involving human subjects and, in the other, not. This apparent inconsistency, based on researcher expertise or qualifications in Internet research, is not unique to Internet research, but is likely to become increasingly relevant as more and more datasets become available. There has been extensive discussion on the IRB Forum (http://www.irbforum.org) on this issue. See, for example, January 11 – January 25, 2012 discussion on "the meaning of anonymity."

Q5: What is *intervention or interaction* with a subject in research on the Internet?

Intervention as defined by the Common Rule "includes both physical procedures by which data are gathered … and manipulations of the subject or the subject's environment that are performed for research purposes."[15] Manipulations of subjects can take many different forms, often mimicking "real-world" manipulations,[16] and manipulation of environments may include testing of different website interfaces, provision of different responses to web queries, recording Internet-based activities or behaviors for subsequent analysis, etc., using the Internet as a reminder or interface for the performance of some physical activity (e.g., taking a medication or performing a task), or may be through something as simple as the presence of a researcher.

Interaction includes "communication or interpersonal contact between investigator and subject."[17] Online interaction may occur in environments that range from virtual worlds or guilds to social media sites to chat rooms, newsgroups, and mobile platforms. Environments can be textual and/or graphical. The interaction itself can include, for example, interviews, focus groups, dialogue across a **listserv** or newsgroup, or any exchange via social media. Surveys presented online should be considered "interaction" with subjects, even if there is no live individual receiving responses in real time but data are collected by the survey engine for later access by the investigator. Interaction can also take place via mobile devices, tablets, or other devices that are connected to Internet applications or tools.

Q6. What are the characteristics of purely public sites?

The analogies of public parks and public libraries have been invoked in Internet research, with the idea that just as there should be no expectation of privacy in real-world public settings, so should some Internet-based settings be considered public. However, questions of access, logging and storage and transmission of data, and other technical considerations complicate the comparisons. Further, just as eavesdropping may not be considered appropriate behavior (even if the activity being observed occurs in a public setting), so too may the monitoring of some Internet-based activities raise similar ethical concerns.

In general, purely public sites fall into one or more of the following categories.

(1) Sites containing information that, by law, is considered "public." In most cases information from these sites will be available without restriction, although access to the information may require payment of a fee. Many federal, state, and local government sites are included in this category: property tax records, birth and death records, real estate transactions, certain court records, voter registration and voting history records, etc.

(2) News, entertainment, classified, and other information-based sites where information is posted

[15] 45 CFR 46.102(f)

[16] See for example, the "Virtual Milgram" at http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0000039.

[17] 45 CFR 46.102(f)

for the purpose of sharing with the public.

(3) Open access data repositories, where information has been legally obtained (with IRB approval if necessary) and is made available with minimal or no restriction.

(4) Discussion fora that are freely accessible to any individual with Internet access, and do not involve terms of access or terms of service that would restrict research use of the information.

Q7: What is *observation of public behavior* online?

If an activity (textual, visual, auditory) is legally available to any Internet user without specific permission or authorization from the individual being observed, or from the entity controlling access to the information, the activity should be considered "public behavior." Examples include "comment" postings on news sites; posting on publicly available hosting sites such as YouTube® or Flickr®; postings on classified sites such as Craigslist®; and postings on unrestricted blog or wiki sites. Information posted on social networking sites such as Facebook®, LinkedIn®, Myspace®, or similar fora, and available without restriction to any authorized user of the site, should also be considered "public behavior," even though access to the website itself may be restricted to individuals who have established an account to use the site.[18] Note that the mere fact of an activity being considered "public behavior" does not mean that observation of the activity should automatically be considered exempt from the requirement of IRB review. Per 45 CFR 46.101(b)(2), if the information is recorded in a way that permits identification of subjects, and if disclosure of the identifiable information could "reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation," then the research would not be exempt from IRB review.[19]

Q8: Is online education *normal educational practice?*

Yes, it often is. The exemption at Section 46.101(b)(1) cites "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." How far beyond the traditional classroom has the widespread use of personal computers and mobile technologies expanded the range of "commonly accepted educational settings"? There are now multiple types of online educational practices ranging from complete degree programs, to individual for-credit classes, to activities that supplement regular classroom instruction, to less formal not-for-credit activities such as instructional videos, online lectures, or TED talks. Considerations include the nature of the "education" being provided; the prevalence of a particular intervention in the learning group under consideration; and the existence of the intervention or teaching process prior to a researcher's involvement. The burden of demonstrating that a particular online educational research

[18] If access to a site is restricted to individuals who must meet specified eligibility criteria, in addition to registering for participation (for instance, individuals who suffer from a particular medical condition), activity on the site should not be considered "public behavior."

[19] The implication of 101(b)(2)—" information obtained is recorded in such a manner that human subjects can be identified…"—is that the information is recorded **by the investigator**. Confirmation of this interpretation would be helpful.

activity should be exempt from IRB oversight may have to rest with the investigator, but IRBs should understand that the range of Internet-enabled "normal educational practices" continues to broaden.

Q9: When is information *recorded in identifiable manner?*[20]

The exemption at 101(b)(2) refers to "Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation."Is "recorded" in this context limited to data that the investigator records him/herself? We believe the intent of this section is "recorded *by the investigator* in such manner…." (See footnotes 20-21).

Q10: When are data, documents, or records *publicly available* on the Internet?

Publicly available may mean:

- Available at no charge to anybody with a computer
- Available to anybody willing to pay the requisite fee
- Available to anybody who meets the terms of a use agreement

Documents that used to be housed in public courthouses or agencies are now often available in electronic form. Such records as state agency reports, property tax assessments, marriage licenses, real estate transactions, voter registration, and the like are now searchable online. Internet tools and sites have simply made access to such public documents easier, but the essential nature of the data is still public.

With the growing availability of data banks and data repositories, and with established data sharing mandates, investigators have greater access to data and data sets. Many IRBs have established exemptions for data shared through ICPSR**,** NIH, NSF, the US Census, etc. Research involving publicly available datasets, with or without identifiers, does not require IRB review under 45 CFR 46.

We may consider these criteria from the United Kingdom's Data Archive, for example,[21] which controls investigator access and the extent to which they are "publicly available":

"For confidential data, the Archive, in discussion with the data owner, may impose additional access controls which can be:

- needing specific authorization from the data owner to access data

[20] 101(b)(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

[21] http://www.data-archive.ac.uk/create-manage/consent-ethics/access-control

- placing confidential data under embargo for a given period of time until confidentiality is no longer pertinent
- providing access to approved researchers only
- providing secure access to data by enabling remote analysis of confidential data but excluding the ability to download data."

Each of the above bullets would constitute a limitation on public availability and the first three would often preclude applicability of the exemption at 46.101(b)(4).

There have been situations where data are under the control of an individual (or entity) who is unaware of regulatory or statutory restrictions on the sharing of data, or who is aware of the restrictions but nevertheless makes the data available (~~i~~ncidents involving WikiLeaks, for example). Investigators and IRBs should ensure that data represented as "publicly available" are, indeed, available without restriction that would limit the proposed use.

Q11: How do investigators obtain the informed consent/parental permission/assent of subjects for research on the Internet?

As with other forms of research, the consent/assent process for Internet research should be tailored to the risks and complexities of the research. The absence of direct, in-person contact can add complications to the consent process.

Three oft-cited concerns with consent in Internet research are verifying identification, ensuring comprehension, and obtaining appropriate documentation when needed. Adequate identity verification may in some cases be handled by the hosting survey provider; in other cases, with minimal risk research, it may not be a critical issue. (See Q 15, below.) Comprehension of the consent materials may be addressed by a checkbox ("I understand and agree") for low risk research, or by mandatory quizzes as a comprehension check. The federal ESIGN law authorizes electronic signatures in certain contexts. In other contexts, state law may control. (Note: OHRP is currently working on issues of e-signatures; see Q14)

The consent process for non-exempt online surveys may include a statement that the subject gives evidence of agreement to participate in the research by the fact of his/her completing the survey. This is permissible even if the consent document does not include all the elements prescribed at 45 CFR 46.116(a), so long as the IRB approves a waiver or alteration of some elements under 46.116(c) and (d).[22]

When obtaining consent in more than minimal risk research, many steps may be necessary. In one example, an industry-sponsored online Phase IV clinical trial, subjects informed of and interested in participation had to meet eligibility criteria; those who qualified underwent ID/age verification. Consent documents were then emailed, faxed, or made available on a web site, with additional information provided in audio or video format. Subjects were required to take a comprehension quiz after reviewing the consent materials and had to score 100% to move ahead. A designated contact for questions and to provide additional information was available to subjects at all times. Applications

[22] See http://answers.hhs.gov/ohrp/questions/7249

such as Skype® or LiveChat® have also been used to enable direct communication between researcher and subject during the consent process.

Research with minors raises particular concerns. There are age verification software products available, which may be of use to researchers. Verification of age can take place through less formal fact-checking embedded in the research instruments (for instance, cross-validating multiple age and birth date questions). Researchers may advertise only on sites that are age-limited to begin with. Coordinating parental consent with child assent can be difficult, and the Children's Online Privacy and Protection Act (COPPA) mandates parental permission if subjects under the age of 13 are being recruited and they provide identifiable information.[23]

Where the HIPAA Privacy Rule applies, the Rule allows a HIPAA authorization for research to be obtained and signed electronically, provided any electronic signature is valid under applicable law.

Q12: When may investigators seek to waive or alter the informed consent of subjects in research on the Internet?[24]

Per 46.116 (c) and (d), to waive some or all of the required elements, or to waive the requirement for consent *in toto,* the IRB must find and document:

The research presents no more than minimal risk; the waiver will not adversely affect subjects' rights and welfare; the research could not practicably be carried out without waiver; and, when appropriate, subjects will be provided with additional pertinent information after participation.[25]

In the absence of a robust identity verification process, some IRBs will only approve an online consent process under circumstances that meet the criteria for alteration/waiver at 46.116(d), and will consider the age/identity verification difficulties as key to the "impracticability" determination. However, if identifiable information may be collected about children under the age of 13, the COPPA requirement for parental consent will apply (there is no waiver provision).

Q13: How do investigators document the informed consent of subjects for research on the Internet?

According to HHS/OHRP,[26] "For most research, informed consent is documented using a written document that provides key information regarding the research. The consent form is intended, in part,

[23] See http://www.coppa.org/comply.htm; also see FTC's August 1, 2012 proposed changes to COPPA, which includes changes to the COPPA definition of "'person information' to include persistent identifiers" (Ropes & Gray, 2012).

[24] See Subpart A Subcommittee, SACHRP Recommendations Regarding the Provisions for Waiver or Alteration of the Informed Consent Requirements Under Department of Health and

Human Services (HHS) Regulations at 45 CFR 46.116(d) (www.hhs.gov/ohrp/archive/.../WaiverConsentDocSAS.doc)

[25] Similar conditions apply to the alteration or waiver of a HIPAA authorization in the research context. Under the Privacy Rule, before approving a waiver or alteration, an IRB or Privacy Board must determine that the use or disclosure of PHI for the proposed research involves no more than minimal risk to individuals' privacy; the research could not practicably be conducted without the waiver or alteration; and the research could not practicably be conducted without access to and use of the PHI.

[26] See http://answers.hhs.gov/ohrp/categories/1566

to provide information for the potential subject's current and future reference and to document the interaction between the subject and the investigator. However, even if a signed consent form is required, it alone does not constitute an adequate consent process. The informed consent process is an ongoing exchange of information between the investigator and the subject and could include, for example, use of question and answer sessions, community meetings, and videotape presentations. In all circumstances, however, individuals should be provided with an opportunity to have their questions and concerns addressed on an individual basis."

Appropriate methods for documenting of informed consent in Internet-based research should reflect the risk and complexity of the research. For straightforward minimal risk research, documentation might be in the form of a simple click-through "I agree" statement preceding access to the study materials, where subjects are presented the appropriate consent information and then signal their consent either by checkbox or by completing the survey or experimental materials; this is consistent with OHRP guidance for survey research.[27] When the research protocol is more complicated, or may present more than minimal risk, a signed document, sent via traditional methods or completed via e-signature (see Q 14) may be a necessary component. Investigators can discuss the informed consent process via chat, email, video, or other online venue, such as in a virtual world. Verification of comprehension can be a challenge. Studies may include a "quiz" or survey after the subject reads or listens to the consent script, to confirm their understanding of the presented materials, and only after completion of the comprehension check will a subject proceed to the study site. Other possibilities include a designated chat room or email contact to discuss the consent process and to allow investigators and participants to converse prior to beginning the research.

Also see Q11, Q12 above.

Q14: Can an electronic signature be used to document consent or parental permission?

This question has been answered at http://answers.hhs.gov/ohrp/questions/7249. "Yes, under certain circumstances. First, the investigator and the IRB need to be aware of relevant laws pertaining to electronic signatures in the jurisdiction where the research is going to be conducted.

"Unless the IRB waives the requirement for the investigator to obtain a signed consent or permission form based on the HHS regulations at 45 CFR 46.117(c), a written consent or permission form, which may be an electronic version, must be given to and signed by the subjects or the subjects' legally authorized representatives or the parents of subjects who are children. Some form of the consent document must be made available to the subjects or the parents of subjects who are children in a format they can retain. OHRP would allow electronic signature of the document if such signatures are legally valid within the jurisdiction where the research is to be conducted.

"OHRP does not mandate a specific method of electronic signature. Rather, OHRP permits IRBs to adopt such technologies for use as long as the IRB has considered applicable issues such as how the electronic signature is being created, if the signature can be shown to be legitimate, and if the consent or permission document can be produced in hard copy for review by the potential subject. One method of allowable electronic signatures in some jurisdictions is the use of a secure system for electronic or

[27] See http://answers.hhs.gov/ohrp/questions/7249

digital signature that provides an encrypted identifiable "signature." If properly obtained, an electronic signature can be considered an "original" for the purposes of recordkeeping."

The HIPAA Privacy Rule also allows HIPAA authorizations to be obtained electronically from individuals, provided any electronic signature is valid under applicable law. In addition, FDA has issued guidance regarding electronic signatures and records at 21 CFR Part 11, with updates in 2007,[28] with specific attention to audit trails and monitoring of research and adverse events. Trace data, logs, time stamps, and electronic data capture can be used.[29]

Q15: Are investigators required to confirm the real identities of the subjects of their Internet research?

Investigators and IRBs should be aware that identity verification is a major issue in Internet research. Absent appropriate verification of a subject's identity, data validity and reliability may be questioned. The need for identity confirmation should take into account:

(a) the importance to the research (i.e., are there critical eligibility criteria? Is there a likelihood of repeat or fraudulent participation, whether for mischief or to collect multiple payments?)

(b) the level of risk to subjects. Low-risk surveys where parental consent could be waived may require only minimal identity verification, perhaps a checkbox. High-risk studies involving the transmission of sensitive information may warrant multiple-factor authentication, such as passwords delivered by mail or telephone, or via an identity verification software or vendor.

(c) There may be a third-party policy or terms of agreement in place that the researcher should acknowledge when considering identity confirmation. For example, Facebook® has a "real-name" only policy, so anonymity is not possible. The norms and expectations of users and venues must be considered.

Online clinical trials, in particular, may include the need for in-person identity verification. Legal jurisdiction should be considered (see Q17). Some IRBs have published suggestions for subject authentication, ranging from sophisticated technical measures such as electronic key exchanges, to less technical personal identification numbers.[30]

Q16: How does legal jurisdiction apply in Internet research?

[28] See http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072322.pdf

[29] See for example a level 1 clinical trial use of electronic monitoring: Internet-based technology facilitates clinical outcome data collection and adverse event monitoring, Orthopedics Today, February 2012 (http://www.healio.com/orthopedics/business-of-orthopedics/news/print/orthopedics-today/%7B523CD70A-7C15-4B5D-8E54-923E8BF958EE%7D/Internet-based-technology-facilitates-clinical-outcome-data-collectionand-adverse-event-monitoring)

[30] See the Pennsylvania State University's statement regarding recruitment for Internet research: "Investigators are advised that authentication - that is, proper qualification and/or identification of respondents - is a major challenge in computer- and internet-based research and one that threatens the integrity of research samples and the validity of research results. Researchers are advised to take steps to authenticate participants. For example, investigators can provide each study participant (in person or by U.S. Postal Service mail) with a Personal Identification Number (PIN) to be used for authentication in subsequent computer- and internet- based data collection." (http://www.research.psu.edu/policies/research-protections/irb/irb-guideline-10)

Jurisdictional authority is complicated by the dispersed nature of Internet subjects and participants. In general, IRBs may assume the jurisdiction of the researcher, not the participants, is controlling. However, in telemedicine, the precedent has held that the jurisdiction of authority is the location of the subjects/patients, consistent with laws regarding the practice and distribution of medicine. This can highlight state and international differences in law and policy. With online clinical trials, for example, state regulations may prevent the enrollment of subjects unless the Investigator is licensed to practice medicine in the state(s) from which subjects are drawn.

If data are stored in "the cloud," (i.e., at multiple, dispersed sites) additional considerations, including data privacy laws at the local storage site(s) and regulations other than those at the research site, may apply. These include, for example, the European Data Privacy Directive 95/46EC or the Canadian *Privacy Act* and the *Personal Information Protection and Electronic Documents Act.* Agreements with data storage and processing entities should acknowledge the investigator's and any business associates' responsibilities to comply with relevant requirements, and subjects should be informed of such arrangements as appropriate.[31]

Q17: What is minimal risk in Internet research?

102(i)[32]: Minimal risk: The regulatory definition of "minimal risk" predates use of the Internet as a communications and research tool. Many Internet-related risks such as identity theft, other types of electronic fraud or security breaches, online "addictions," and electronic monitoring, stalking, or bullying, can have serious consequences, but most were not part of our daily lives—or indeed even contemplated—when the regulations were first written. It is increasingly appropriate to include the risk of computer-related harms, such as hacking, phishing, breach, lack of appropriate security measures, etc., as among those risks encountered in daily life.

As with any form of human subject research, there runs a continuum of risk in Internet research, and the type of IRB review—expedited or full board—should reflect the level of anticipated risk. Categorization schedules such as the University of North Carolina's Data Security Recommendations[33]or the Harvard Research Data Security Policy[34] can help IRBs determine appropriate protections for data of differing levels of sensitivity. Use agreements may be necessary supplements to protocols, especially those in cross-agency, cross-institutional, or multi-site studies. In addition, where appropriate under NIH standards, Certificates of Confidentiality offer protection

[31] For example, many Canadian research ethics boards include the statement "Please note that the online survey is hosted by "Survey Monkey" 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 U.S. Patriot Act that allows authorities access to the records of internet service providers. If you choose to participate in the survey you understand that your responses to the questions will be stored and accessed in the USA. The security and privacy policy for Survey Monkey can be viewed at http://www.surveymonkey.com/" on their consent documents.

[32] 45 CFR 46.102(i): *Minimal risk* means that the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.

[33] See http://research.unc.edu/ccm/groups/public/@research/@hre/documents/content/ccm3_035154.pdf

[34] See http://security.harvard.edu/research-data-security-policy

against compelled disclosure.[35]

Investigators and IRBs must consider both risks related to the specific research protocol and risks related to the technologies in use. How should IRBs think about these risks, and how can they accurately be conveyed to subjects—especially when the full extent of risks might not be known even to the investigator? While subjects may be reassured by being told that appropriate precautions will be taken to ensure the security of their data, the exact nature of "appropriate precautions" (in the absence of published guidelines or established standards) can be difficult to determine or to convey in a meaningful way. IRBs that regularly review Internet research should consider including one or more information technology professionals on their rosters, to assist with determinations of actual risk and to advise on implementation of appropriate security measures.

Since research conducted online may not involve real-time communication with participants, misunderstanding or distress may not be evident to the researcher, which can in turn elevate the risk to subjects. IRBs must be aware of these possibilities; some types of Internet research may not be approvable without assurance that immediate contact with a researcher will be available if necessary.

Q18: How may investigators minimize risk of harm when using sensitive online data?

The definition of minimal risk references both the *probability* and the *magnitude* of harm, and investigators and IRBs must consider both dimensions. A risk of significant harm (e.g., identity theft, breach of confidential medical or personal information) that is technically possible, but of small likelihood, may be judged to be minimal if the IRB is satisfied that the investigator's data security procedures are consistent with best-practice recommendations of the institution's IT professionals. Common guidance, or reference to established standards, would be helpful.[36]

Sensitive data include personal health, economic, educational, and/or reputational information, and may be more readily available in online venues than in traditional onground research. IRBs should consider changing sections on consent forms from "Confidentiality" to "Limits to Confidentiality" and should ensure accurate use of terms such as "anonymous" and "confidential."

While the HIPAA standards for protection of data may be extreme for much social/behavioral/educational research, consistent standards for low to minimal risk research should include consideration of how subjects' data will be collected, transmitted to the researcher, and stored. If a third party venue or processing site is involved, their access to and storage of those data should be specified. The consent process should include explanations of how data are maintained, ranging from individually identifiable forms to aggregate forms, and what linking or reidentification measures are

[35] See NIH, "any research project that collects personally identifiable, sensitive information and that has been approved by an IRB operating under either an approved Federal-Wide Assurance issued by the Office of Human Research Protections or the approval of the Food and Drug Administration is eligible for a Certificate. Federal funding is not a prerequisite for an NIH-issued Certificate, but the subject matter of the study must fall within a mission area of the National Institutes of Health, including its Institutes, Centers and the National Library of Medicine." (http://grants.nih.gov/grants/policy/coc/faqs.htm#278)

[36] See, for example, http://cphs.berkeley.edu/datasecurity.html and http://security.harvard.edu/research-data-security-policy

possible. If aggregated anonymized data will be made publicly available, investigators and IRBs should consider whether subjects could be (re)identified and how that likelihood could be minimized.

Whenever possible, identifiable data should be encrypted in transit (for most low to minimal risk studies, basic SSL encryption is acceptable) and while at rest (whole disk encryption is readily available). Data should be unlinked from identifiers and IDs destroyed as soon as they are no longer needed. Researchers should consider provisions for remote locking of devices or remote destruction of data in the event of a lost device. When investigators are entrusted with data and devices, they have a responsibility to minimize risk and to honor their obligations to subjects.

Both data use and data management plans should reflect investigators' and subjects' responsibilities and rights, including any applicable privacy laws and regulations. In addition to standard elements regarding access, longevity, and ownership of data, plans should identify available resources in the event of harms.

Social media sites, search engines, and virtually all online fora retain log data. A shared knowledge base of appropriate characteristics of venues and tools for IRBs and researchers would be helpful to understand the data life cycle on the most commonly used online research sites and tools.

Q19: What forms of online recruitment are used and what is reviewable by an IRB?

Recruitment tools include Web ads, Twitter streams, blog postings, YouTube videos, and "push" methods, such as email solicitations and texts. Links to online recruitment sites (e.g., Patients Like Me, Inspire) may also be provided in other media (television, newspaper, classified, public transit posters, robo-calls, etc.)**.** OHRP considers direct subject recruitment part of informed consent, which is subject to IRB review.

Note that, per FDA guidance, prior IRB review is not necessary for simple listings of clinical trials on websites where the system format limits the provided information to basic descriptive information, including study title, purpose of the study, protocol summary, basic eligibility criteria, study site location(s), and how to contact the study site for further information.[37]Any recruitment plan must receive IRB review and approval prior to initiation if additional information is provided, including description of research risks, potential benefits, incentives (monetary or non-monetary), or where identifiable information is solicited to determine eligibility.

As with other forms of research, introducing an investigator into a forum or study site may be

[37] Specifically, refer to clinicaltrials.gov, where study listings that meet the posting criteria of the site need not, in and of themselves, be reviewed and approved by an IRB prior to posting. However,

"Most trials require approval from a human subjects review board. If your study requires approval, you may register your study on ClinicalTrials.gov prior to getting approval if the Overall Recruitment Status of the study is "Not yet recruiting." See Overall Recruitment Status data element on ClinicalTrials.gov .

If a study requires human subjects review board approval, approval must be obtained before the study's Overall Recruitment Status is changed to Recruiting. When board approval is obtained, please update the protocol section of the study record in the Protocol Registration System (PRS) and release the study for processing." (http://clinicaltrials.gov/ct2/manage-recs/faq#board)

appropriate, and the investigator and IRB should review the introduction process as part of the recruitment plan. A moderator, high-ranking member, or other member of status can provide information to the online site community prior to the researcher's entrance.

Q20: How is deception conducted in Internet research?

Occasionally some aspects of a study are not fully disclosed in advance, to avoid affecting subjects' responses. Deception in Internet research may be ethically complex; Kraut *et al.* (2003) note greater difficulty in monitoring adverse effects and in provision of adequate debriefing.[38] Internet research provides many opportunities for deception.[39] Researchers can create "fake" or alternative locales to observe behavior or actions; provide limited or erroneous information to see how subjects respond to a given situation;[40] or send "spam" or "phishing" messages to elicit personal data.[41]Because informed consent in a deceptive study is necessarily limited, the need for appropriate debriefing following participation must be given special consideration, but difficulties abound. Subjects may choose to leave a venue or locale without reading (or even seeing) the debriefing material; may change email addresses; or may fail to respond to electronic communications. Investigators and IRBs should be aware of these potential challenges when considering appropriate debriefing measures.[42]

[38]SeeKraut, R. *et al.* (2003). Psychological Research Online: Opportunities and Challenges. http://www.apa.org/science/leadership/bsa/internet/internet-report.pdf

[39] See Bachard, K. and Williams, J. (2008). Practical advice for conducting ethical online experiments and questionnaires for United States psychologists.Behav Res Methods. ;40(4):1111-28.

[40] See for example the Virtual Milgram experiment at http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0000039

[41] See Finn, P. and Jakobsson, M. (2007). Designing and Conducting Phishing Experiments. IEEE. http://markus-jakobsson.com/papers/jakobsson-ieeets07.pdf

[42] See for example, University of Massachusetts: "Some research requires a debriefing after participants have completed an online survey. Online debriefing forms should be similar to the debriefing process done during in-lab experiments. The debriefing page should come immediately after the last question on the survey. Participants should be thanked for participation and more information as to the purpose of the study should be provided. Also, researchers contact information and information about other resources (IRB info, Health Services, Local Resources) should be provided and participants should be reminded to print a copy of the debriefing form for their records. Participants should also be given the option to withdraw their data at this point (now that they have been fully informed as to the intent and purpose of the study). If they agree to have their data used for the study then they should have an "I Agree" button to click and submit their data online. If they do not agree to have their data used in the study they should have an "I Do Not Agree" button to click so that their data is not submitted and collected online. Please check with the online survey program you are using to ensure that these capabilities are allowed." (http://www.umass.edu/research/online-surveysurvey-research-guidance)