REGULATING HUMAN SUBJECTS RESEARCH IN THE INFORMATION AGE: DATA MINING ON SOCIAL NETWORKING SITES

*Lauren B. Solberg**

In the 1970s, the Department of Health, Education & Welfare promulgated the first regulations governing research with human subjects. Currently, the Department of Health and Human Services regulates both biomedical and behavioral research with human subjects through 45 C.F.R. § 46, but these regulations have not been revised to address the advances in technology that have changed the ways in which human subjects research is conducted in the 21st century. One of these advances includes researchers' use of the Internet for various research purposes, including recruiting subjects for their studies, as well as mining data on social networking sites. The issue of data mining on social networking sites for research purposes is becoming more important as researchers in a number of academic fields are collecting research data in this manner. The question of whether such research involves "human subjects" as the term is defined in 45 C.F.R. §46 is important because the answer dictates whether or not such research is subject to federal regulation and oversight.

The U.S. Department of Health and Human Services' Office for Human Research Protections (OHRP) has previously acknowledged an increased use of the Internet for research purposes, but it has not published any guidance about how the Office understands the issue (nor have any applicable specific regulations been promulgated). Such guidance, and particularly guidance that addresses the use of social networking sites for research purposes, is becoming increasingly more important as researchers are expanding beyond just Facebook and MySpace to recruit research participants and collect data. Furthermore, new social networking sites such as PatientsLikeMe.com are changing the ways in which both behavioral and biomedical researchers recruit and interact with study participants.

This Article identifies three key research activities that any new guidance that the OHRP publishes about Internet research should address: (1) the use of the Internet (including social networking sites) to recruit subjects; (2) the use of the Internet to collect personal information via direct interaction with site users; and (3) the collection of personal information from the Internet for research purposes without direct interaction with the owner of the information, i.e. data mining. It further explores recent case law

* Assistant Vice President of Human Subjects Protection & Research Integrity, Meharry Medical College, Nashville, TN. J.D., Vanderbilt University; M.T.S., Harvard University; B.A., University of Florida. The author is grateful for the thoughtful comments and questions from the participants of the Northern Kentucky University Law Review Law & Informatics Symposium which contributed to the final version of this article. The author also wishes to acknowledge the editorial assistance of Adam Shniderman with this article. This work was supported in part by Meharry R-Center grant 5 U54 RR026140-03(NCRR)/8 U54 MD007593-03 (NIMHD), *Vanderbilt CTSA* grant 5UL1 RR024975 (NCRR), and 3S21MD000104-11S1 (NIH/NIMHD).

that says that there is no reasonable expectation of privacy with respect to information posted on social networking sites, and discusses the history of privacy in the federal regulations governing research with human subjects. This Article concludes that in light of the changes in online social networking and resulting changes in expectations of privacy, as well as the consideration that the Department of Health and Human Services is giving to revising 45 C.F.R. § 46, guidance – or even new regulations – should be drafted that address these key research activities. It furthermore suggests language that would inform researchers and research institutions whether each of these key research activities is subject to the regulations governing research with human subjects.

I. INTRODUCTION

The Internet, and specifically the prevalence of social networking sites like Facebook, Twitter, and MySpace, has greatly changed the landscape of human subjects research.[1] When 45 C.F.R. § 46, the federal regulations that govern the conduct of research with human subjects, was first promulgated in 1974, the Department of Health, Education & Welfare (HEW) could not have anticipated the ways in which the use of a computer would affect the conduct of human subjects research in the future. At that time, protecting an individual's private information that might be revealed during the course of research appeared only to be of general concern to HEW.[2] Instead, media attention was focused on the revelation to the American public of the U.S. Public Health Service study of syphilis in Tuskegee, Alabama, often referred to as the Tuskegee syphilis study.[3] Between 1932 and 1972, approximately 400 African-American men in Alabama with syphilis were left untreated by U.S. government researchers who wanted to understand the course of the disease.[4] The regulations were promulgated in large part in response to this atrocity, and were initially intended to protect individuals against unnecessary risk that they might incur during the course of their participation in biomedical research.[5]

The regulations have evolved over time to seek to protect against any number of risks that participants in biomedical as well as in behavioral and

1. *See infra* Part III.

2. Contrary to today, where a "human subject" is – among other things – one about whom a researcher obtains identifiable private information, no mention of privacy exists in the 1974 regulations. *See* 45 C.F.R. §46.102(f) (defining a "human subject" in research); *infra* Part II discussing the history of the regulations governing human subjects research.

3. U.S. Public Health Service Syphilis Study at Tuskegee: The Tuskegee Timeline, CENTERS FOR DISEASE CONTROL AND PREVENTION, http://www.cdc.gov/tuskegee/timeline.htm (last visited Aug. 1, 2012).

4. Jean Heller, *Syphilis Victims in U.S. Study Went Untreated for 40 Years*, N.Y. TIMES, July 26, 1972, at 1, 8.

5. *See* Secretary's Interpretation of "Subject at Risk," 41 Fed. Reg. 26572 (June 28, 1976) (explaining that "[t]he regulations were not, and have never been, intended to protect individuals against the effects of research and development activities directed at social or economic changes. . . "). *See also infra* Part II for a discussion of the history of the regulations governing human subjects research.

social science research may incur, and have evolved to provide special protections for vulnerable populations in research.[6] However, the last time that Subpart A, the provisions in 45 C.F.R. § 46 that govern research with all human subjects, was revised was in 1991.[7] The dated nature of these regulations necessarily means that they specifically do not take into account developments in technology, such as the Internet, that have changed both people's perceptions of privacy as well as the general conduct of human subjects research.[8]

The fact that changing perceptions of privacy in conjunction with recent technological advancements are not accounted for in 45 C.F.R. § 46 often makes it difficult to interpret the regulations as currently written, and in particular makes it difficult to determine when human subjects are involved in research when that research involves the use of the Internet. A human subject is "a living individual about whom an investigator (whether professional or student) conducting research obtains (1) [d]ata through intervention or interaction with the individual, or (2) [i]dentifiable private information."[9] Private information is identified as "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)."[10] Because the identification of research as "human subjects research" may hinge on whether private information is being obtained about a living individual, it is important for researchers and Institutional Review Boards (IRBs),[11] committees charged by 45 C.F.R. § 46 to review research with human subjects, to know when private information is being obtained. The ability

6. *See, e.g.*, Additional Protections for Children Involved as Subjects in Research, 43 Fed. Reg. 9814 (Mar. 8, 1983) (to be codified at 45 C.F.R. pt. 46); Final Regulations Amending Basic HHS Policy for the Protection of Human Research Subjects, 46 Fed. Reg. 8366 (Jan. 26, 1981) (to be codified at 45 C.F.R. pt. 46); Protection of Human Subjects: Prisoners Used as Subjects in Research, 45 Fed. Reg. 36386 (May 30, 1980) (to be codified at 45 C.F.R. pt. 46); Additional Protections Pertaining to Research, Development, and Related Activities Involving Fetuses, Pregnant Women, and In Vitro Fertilization, 43 Fed. Reg. 1758 (Jan.11, 1978) (to be codified at 45 C.F.R. pt. 46).

7. 45 C.F.R. §46 (2011). While Subpart A pertains to research with all human subjects, Subparts B, C, and D of 45 C.F.R. §46 govern research with pregnant women and fetuses, prisoners, and children, respectively. *Id.* Subpart A is often referred to as the "Common Rule." *See infra* note 194.

8. *See* discussion i*nfra* Part III (describing different ways that the Internet is used in research).

9. 45 C.F.R. §46.102(f) (2012).

10. *Id.*

11. The IRB is a committee typically composed of faculty, staff, and/or students at a research institution that reviews all human subjects research to ensure that it is both ethical and compliant with these federal regulations. *See* 45 C.F.R. § 46.107 (2012) (stating the requirements for IRB membership). Once constituted, the IRB reviews the research to determine whether subject selection is equitable, whether the risk is proportional to the benefits, and whether the confidentiality of any data that are collected will adequately be protected. 45 C.F.R. § 46.111 (2012) (detailing the criteria that the IRB must confirm are satisfied before approving a project).

to make this determination is more difficult in our technological age, and requires an understanding of when individuals have a reasonable (or unreasonable) expectation of privacy.

Unfortunately, the Office for Human Research Protections (OHRP), the office in the Department of Health and Human Services (HHS) (formerly HEW, until 1980) that interprets and enforces 45 C.F.R. § 46, has published almost no guidance or other informal advice, *e.g.,* in the form of a letter or list of FAQs, about the use of the Internet in human subjects research.[12] Without input from HHS through informal rulemaking or other means, researchers and IRBs are left to interpret the current regulations themselves when it comes to respectively designing and reviewing studies that involve the use of the Internet. Because the use of the Internet is prevalent in biomedical, social, and behavioral research studies,[13] there are countless numbers of researchers – not just researchers in one particular field or type of field – who would benefit from new regulations or guidance that clarify when research facilitated by the Internet is human subjects research.

This article focuses in particular on one way in which the Internet is used to conduct research – data mining on social networking sites. With the increasing prevalence of the use of social networking sites, academic researchers have realized the vast amount of data that exist to be analyzed, so predictably they are logging on to these sites and reviewing the information that people post in order to draw various conclusions about human thought and behavior.[14]

While some uses of the Internet are more clearly used to facilitate human subjects research, or are, in and of themselves, human subjects research as defined in 45 C.F.R. § 46,[15] it is not entirely clear whether researchers who collect data for a study solely by mining social networking sites are engaging in research with "human subjects" as defined in 45 C.F.R. § 46.102(f). This is in large part because of changing notions of privacy. My previous work analyzed the definition of a human subject, focusing on whether data mining on social networking sites like Facebook is research with human subjects and, as such, is

12. *See* United States Department of Health and Human Services, Policy and Guidance Index, http://www.hhs.gov/ohrp/policy/index/index.html (last visited Aug. 9, 2012). The only information that the OHRP has made available online regarding Internet research is this guidance document that discusses what IRBs should look for when reviewing advertisements published on clinical trials web sites. *See infra* text accompanying note 63. Although helpful, this single document does not begin to address the numerous other ways in which researchers use the Internet to facilitate their research, e.g. for distribution of surveys, or for visits to chat rooms. *See* discussion of such uses of the Internet to facilitate the conduct of human subjects research *infra* Part III.A..

13. *See infra* Part III.A. for a discussion of the ways the Internet is used for research.

14. *See infra* Part III.B. for a discussion of data mining on social networking sites.

15. 45 C.F.R. § 46.102(f).

federally regulated.[16] Furthermore, my previous work analyzed what type of IRB review would be required if 45 C.F.R § 46 regulates this type of research.[17]

This prior work concluded that data mining on social networking sites likely does not involve a researcher obtaining data about an individual through intervention or interaction.[18] However, this is only true as long as that researcher does not actively communicate with the individual for purposes of collecting data, for example, there is no verbal exchange, either online or in person, between the two parties.[19] Such research is also not human subjects research if the researcher does not "friend" the individual for the sole purpose of collecting research data.[20] Furthermore, and more importantly for the purposes of this article, I concluded that information posted by individuals on social networking sites likely is not private because, under most circumstances, people can no longer have much reasonable expectation of privacy with respect to information they post online, especially on a social networking site.[21] However, the issue of what "private information," as used in 45 C.F.R. § 46, is in light of the technology readily available today is still unresolved.

Part II of this article discusses the history of privacy in the federal regulations that govern human subjects research. Part III describes the ways in which researchers make use of the Internet for their research, and affirms my earlier conclusion that one of these ways – data mining on social networking sites – likely is not human subjects research, primarily because individuals have a very limited expectation of privacy when it comes to information they post online. Part IV examines how courts view individuals' expectations of privacy with respect to the information they post on social networking sites, and Part V identifies a new type of social networking site that is being used for academic research purposes. In light of the static federal regulations governing human subjects research, the technology that is readily available to researchers, and the limited expectations of privacy that people have on the Internet, Part VI suggests language for new regulations or guidance that address Internet research.[22]

---

16. Lauren B. Solberg, *Data Mining on Social Networking Sites: A Free Space for Researchers or an IRB Nightmare?*, 2010 ILL. J.L. TECH. & POL'Y 311. (2010).

17. *Id.*

18. *Id.* at 321, 324.

19. *Id.* at 324.

20. *See id.* at 321, n.70.

21. *Id.* at 328.

22. The United States Food and Drug Administration (FDA) held public hearings in 2009 to explore the possibility of developing policy or guidance on promoting FDA-regulated products on the Internet and through social media. U.S. Public Hearing on Promotion of FDA-Regulated Medical Products Using the Internet and Social Media Tools, Food and Drug Administration, (Nov. 12-13, 2009), http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm184250.htm. The only publication to date is draft guidance published in December 2011 entitled "Responding to Unsolicited Requests for Off-Label Information About Prescription Drugs and Medical Devices" that mentions social media and the Internet, but was not issued for the specific purpose of addressing social media and the Internet. Draft Guidance on Unsolicited Requests for Off-Label Information, EYE ON FDA, http://www.eyeonfda.com/

II. A HISTORY OF PRIVACY IN 45 C.F.R. § 46

Interpreting the meaning of "private information" as defined in 45 C.F.R. § 46.102(f) requires an understanding of what reasonable expectations of privacy are, and in particular, what they are in light of today's technology. What were considered reasonable expectations of privacy at the time of the initial promulgation of 45 C.F.R. § 46 in 1974, or even in 1991 when Subpart A was last revised, may no longer be applicable. Indeed, "privacy must be evaluated in light of the 'modern enterprise and inventions'"[23] and while "[i]t is still possible to protect privacy, []doing so requires that we rethink outdated understandings of the concept."[24] Thus, it is reasonable for IRBs to believe that some research facilitated by the Internet involves the collection of private information; however, we cannot continue to apply the same standards for determining whether that information is private – particularly when it is collected from social networking sites.

As previously mentioned in the Introduction, the current federal regulations governing research with human subjects define a human subject as "a living individual about whom an investigator (whether professional or student) conducting research obtains (1)[d]ata through intervention or interaction with the individual, or (2) [i]dentifiable private information."[25] The regulations then explain that "[p]rivate 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)."[26]

These definitions are found in Subpart A of the regulations, which underwent its most recent revision in 1991; however, the definitions of a human subject and private information have remained static since a 1981 revision of the regulations.[27] An in-depth examination of the evolution of the definition of a human subject in 45 C.F.R. §46 shows how HHS came to seek protection of

---

eye_on_fda/2012/01/draft-guidance-on-unsolicited-requests-for-off-label-information.html (last visited Aug. 9, 2012). However, researchers who conduct social or behavioral research may not find such guidance to be helpful, as FDA guidance and regulations apply to "clinical investigations," which in summary refers to research in which a drug or device, or other FDA-regulated articles are involved. 21 C.F.R. § 56.102(b) (2012). Thus, revisions to Subpart A or OHRP guidance would be more applicable to those conducting research using the Internet.

23. Connie Davis Powell, *"You Already Have Zero Privacy. Get Over it!" Would Warren and Brandeis Argue for Privacy for Social Networking?*, 31 PACE L. REV. 146, 161 (2011) (quoting Samuel Warren and Louis Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 196 (1890)).

24. *Id.* at 163 (quoting Daniel J. Solove, *Do Social Networks Bring the End of Privacy?*, SCI. AM., Sept. 1, 2008, at 104).

25. 45 C.F.R. § 46.102(f).

26. *Id.*

27. *See* Final Regulations Amending Basic HHS Policy for the Protection of Human Research Subjects, 46 Fed. Reg. 8366, 8387 (Jan. 26, 1981) (to be codified at 45 C.F.R. pt. 46) (defining "human subject," which is the same definition that is currently found at 45 C.F.R. § 46.102(f)).

individuals whose identifiable private information is obtained during the course of research, what HHS believed to in fact be private information, and why clarification of what private information is would assist researchers and IRBs in conducting and reviewing research involving the Internet.

When 45 C.F.R. § 46 was first promulgated in May 1974, the regulations reflected the same fundamental purpose that they do now – to protect individuals enrolled in research. However, the notice of proposed rulemaking published in 1973 and the final rule published in the Federal Register in 1974 reveal that the concepts of privacy and private information were not specific concerns of HEW.[28]

While a "human subject" now is defined in the regulations, the 1974 regulations did not contain such a definition.[29] Although HEW's explanation of the final rule discussed the involvement of "human subjects" in research, and although the regulations themselves used the term "human subject," the first set of federal regulations protecting human subjects in research failed to actually define a "human subject."[30] Instead, these regulations only defined a "subject at risk," which was

> any individual who may be exposed to the possibility of injury, including physical, psychological, or social injury, as a consequence of participation as a subject in any research, development, or related activity which departs from the application of those established and accepted methods necessary to meet his needs, or which increases the ordinary risks of daily life, including the recognized risks inherent in a chosen occupation or field of service.[31]

There was no mention of privacy or private information at all in either the notice of proposed rulemaking or in the final rule.[32]

A notice subsequently published in the 1976 Federal Register explained this omission to a certain extent. The notice provided the HEW Secretary's interpretation of a "subject at risk," indicating that the regulations as promulgated in 1974 were designed to protect people against harms suffered in biomedical research, citing harms such as those that could be suffered as a result of experimentation with "FDA-approved drugs for any unapproved purpose; psycho-surgery and other techniques for behavior control currently being developed . . . use of experimental intrauterine devices; [and] biomedical

---

28. *See* Protection of Human Subjects, 39 Fed. Reg. 13914(May 30, 1974) (to be codified at 45 C.F.R. pt. 46); Protection of Human Subjects, 38 Fed. Reg. 27882 (proposed Oct. 9, 1973) (to be codified at 45 C.F.R. pt. 46).

29. *See* Protection of Human Subjects, 39 Fed. Reg. 13914 (May 30, 1974) (to be codified at 45 C.F.R. pt. 46).

30. *See id.*

31. *Id.* at 18917.

32. *See* Protection of Human Subjects, 39 Fed. Reg. at 13914; Protection of Human Subjects, 38 Fed. Reg. at 27882.

research in prison systems . . . ."[33] The Secretary specifically states that "[t]he regulations were not, and have never been, intended to protect individuals against the effects of research and development activities directed at social or economic changes, even though those changes might have an impact upon the individual."[34] Thus, protecting against the disclosure of a person's private information that was learned during the course of research was not a specific goal of the regulations.

It was not until 1978 when the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (the National Commission), in accordance with its charge in the National Research Act to study IRBs, issued a report and recommendations about IRBs that there was any recognition on the part of the government that private information obtained during the course of research may require special protection.[35] The National Commission's report on IRBs included a definitions section, and the current regulatory definition of a human subject was derived from this report.[36] The National Commission defined a human subject as "a person about whom an investigator (professional or, student) conducting scientific research obtains (1) data through intervention or interaction with the person, or (2) identifiable private information."[37] The National Commission went on to elaborate on the meaning of private information, commenting that

> '[p]rivate 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 (*e.g.,* 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.[38]

---

33. Secretary's Interpretation of "Subject at Risk," 41 Fed. Reg. 26572 (June 28, 1976).

34. *Id.*

35. Institutional Review Boards: Report and Recommendations of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 43 Fed. Reg. 56174 (Nov. 30, 1978). This report issued by the National Commission should not be confused with the Belmont Report, one of the most well-known discussions of the ethical principles that should govern research with human subjects. The Belmont Report, which identified the three fundamental principles of research ethics as respect for persons, beneficence, and justice, was not published in the Federal Register until 1979. Belmont Report, 44 Fed. Reg. 23192 (Apr. 18, 1979).

36. *Id.*

37. Institutional Review Boards: Report and Recommendations of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 43 Fed. Reg. at 56175.

38. *Id.*

The language that the National Commission used to explain private information is now included in 45 C.F.R. § 46.[39] Thus, the only differences between the National Commission's definition of a human subject and the current definition in 45 C.F.R. § 46 are that the current regulations define a human subject as a "living individual," and the current regulations do not limit the scope of the definition to "scientific research," instead seeking to protect participants in simply "research."[40]

Following the National Commission's publication of their report on IRBs, HEW's proposed amendments to 45 C.F.R. § 46 sought to include a definition of a "human subject" for the first time rather than simply a "subject at risk," [41] though the definition at that time did not as closely resemble the definition that the National Commission provided in its report on IRBs.[42] HEW instead proposed to only define a human subject as "an individual about whom an investigator (whether professional or student) conducting research obtains (1) data through intervention or interaction with the person, or (2) identifiable information."[43] Thus, the proposed amendments excluded the term "private" from the definition of a human subject even though the National Commission had expressed concerns about IRBs reviewing the ways in which private information might be used in research.[44] Additionally, a final rule was published in April 1979 allowing for the issuance of certificates of confidentiality under the Public Health Service Act to shield the identity of individuals enrolled in mental health research (further emphasizing the importance of protecting privacy and confidentiality in research).[45]

The HHS explanations provided in the publication of the final rule in 1981 stated that there were fewer than twenty comments submitted to HHS about the proposed definition of a human subject, with some objecting to the broad nature of the definition, indicating that the protection of subjects of non-biomedical research (*i.e.,* subjects of social or behavioral research) under the regulations

---

39. *See* 45 C.F.R. § 46.102(f).

40. *Compare* 45 C.F.R. § 46 (2012) *with* Institutional Review Boards: Report and Recommendations of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 43 Fed. Reg. at 56175.

41. Proposed Regulations Amending Basic HEW Policy for Protection of Human Research Subjects, 44 Fed. Reg. 47688, 47693 (Aug. 14, 1979) (to be codified at 45 C.F.R. §46).

42. *Compare id. with* Institutional Review Boards: Report and Recommendations of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 43 Fed. Reg. at 56175.

43. Proposed Regulations Amending Basic HEW Policy for Protection of Human Research Subjects, 44 Fed. Reg. at 47693.

44. Institutional Review Boards: Report and Recommendations of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 43 Fed. Reg. at 56181.

45. Protection of Identity – Research Subjects, 44 Fed. Reg. 20382, 20382 (Apr. 4, 1979) (to be codified at 42 C.F.R. pt. 2a).

would be unnecessary.[46] However, in an effort to comport with recommendations proffered by the National Commission, the agency chose to adopt the National Commission's definition of a human subject, making only a few minor changes,[47] and this definition is the one included in the regulations today.[48] By including the term "private" in the definition of a human subject, HHS anticipated that the definition would "exempt[] from the regulations nearly all library-based political, literary and historical research, as well as purely observational research in most public contexts, such as behavior on the streets or in crowds."[49]

Of course, HHS could not have conceived of the Internet as a commonly used research tool in academia in the 1970s or early 1980s, much less anticipated it as a means of data collection. However, the shift in HHS's thinking about risk to participants in research, and its resulting shift in its view of who really requires protection as research subjects is an interesting one. HHS took seriously the recommendations of the National Commission in an effort to ensure better protections for human subjects in research.[50] However, the agency has not revised its definition of a human subject or elaborated on the notion of identifiable private information since the 1981 final rule was published.[51] Furthermore, HHS has not shared with IRBs or researchers how it views the notion of privacy in light of the ways that researchers are using the Internet to facilitate their research, and more specifically, have not indicated whether data mining on social networking sites is human subjects research.[52]

As a result of the changing nature of human subjects research, HHS should again consider whether its definition of a human subject is adequate to ensure the protection of the individuals who enroll in research, whether biomedical, social, or behavioral. In particular, HHS should consider whether clarifications regarding what constitutes private information would better protect human subjects and help researchers and IRBs understand to what type of research these regulations are in fact applicable; that is, when researchers are in fact obtaining identifiable *private* information.

---

46. Final Regulations Amending Basic HHS Policy for the Protection of Human Research Subjects, 46 Fed. Reg. 8366, 8373 (Jan. 26, 1981) (to be codified at 45 C.F.R. pt. 46).

47. *Id.*

48. *See* 45 C.F.R. § 46.102(f).

49. Final Regulations Amending Basic HHS Policy for the Protection of Human Research Subjects, 46 Fed. Reg. at 8373.

50. *Id*. at 8366.

51. *Id*.

52. *See generally id.*

III. RESEARCHERS' USE OF THE INTERNET AND INDIVIDUALS' EXPECTATION OF PRIVACY

The technological developments made at the end of the 20th century and beginning of the 21st century have significantly influenced the ways in which human subjects research has changed since the federal regulations governing such research were first promulgated. The Internet in particular has greatly influenced the conduct of biomedical research, as well as research in the social and behavioral sciences. There are three important ways in which the Internet is currently used to facilitate human subjects research: use of online surveys, recruitment of participants for studies through online advertisements, and observation of real-time online behavior of individuals.

A. Ways That the Internet Is Used for Research

Through the use of web sites like SurveyMonkey[53] and REDCap,[54] researchers are now able to administer surveys online to vast numbers of research participants. For example, researchers at the University of Utah sent a survey through SurveyMonkey to directors of burn centers across the United States to analyze the frequency of the use of telemedicine for burn care.[55] University of Kentucky researchers used REDCap Survey to ask orthodontists about the types of retainers they give their patients.[56] The distribution of online surveys such as these is typically considered human subjects research, and therefore requires IRB review.[57]

---

53. SurveyMonkey is a web site that allows users to create and distribute surveys online. SURVEYMONKEY, www.surveymonkey.com (last visited Aug. 9, 2012). Users can create short surveys for a limited number of respondents free of charge, and can pay monthly fees for the right to create more in-depth, more sophisticated surveys. SurveyMonkey Plans & Pricing, www.surveymonkey.com/pricing (last visited Aug. 9, 2012).

54. REDCap (Research Electronic Data Capture) is "a secure, web-based application designed to support data capture for research studies, providing: 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources." REDCap, www.project-redcap.org (last visited Aug. 9, 2012). *See also* Paul A. Harris, Robert Taylor, Robert Thielke, Jonathon Payne, Nathaniel Gonzalez, Jose G. Conde, *Research Electronic Data Capture (REDCap) - A Metadata-Driven Methodology and Workflow Process for Providing Translational Research Informatics Support*, 42 J. BIOMED. INFORM. 377 (2009). In addition to providing a data management system, the program also includes a survey feature that allows for the administration of online surveys.

55. Brennan Holt et al., *Telemedicine Use Among Burn Centers in the United States: A Survey*, 33 J. BURN CARE & RES. 157 (2012).

56. Michael C. Pratt et al., *Evaluation of Retention Protocols Among Members of the American Association of Orthodontists in the United States*, 140 AM. J. ORTHODONTICS & DENTOFACIAL ORTHOPEDICS 520 (2011).

57. The distribution of a survey for academic research purposes is human subjects research as long as the project satisfies the regulatory definition of "research," which is "a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge." 45 C.F.R. § 46.102(d) (2012). The researchers at the University of Utah stated that their study was approved by the IRB. Holt et al., *supra* note 55, at

Additionally, web sites are often used to advertise studies to recruit potential participants. Pharmaceutical company web sites often include information about the clinical trials they are sponsoring, and even give potential participants the opportunity to do a basic assessment of whether or not they may be eligible to participate in a particular clinical trial.[58] Some researchers post advertisements on sites like Craigslist, including information about the study and how to contact the researchers if individuals may be interested in participating.[59] In 2000, the U.S. National Institutes of Health (NIH) established a web site called ClinicalTrials.gov that makes a list of clinical trials being conducted in the United States publicly available, and also includes the results of clinical trials that have since closed.[60]

The advertisements placed on web sites like Craigslist are included with researchers' IRB submissions, and are reviewed along with the protocol (description of study procedures, including method for data collection, recruitment and consent processes, and other applicable information), consent documents, and other written materials.[61] An IRB's review of such materials is justified because recruitment represents the beginning of the informed consent process,[62] which is regulated under 45 C.F.R. § 46.116. Furthermore, the OHRP published guidance in 2005 that stated that "[w]hen information posted on a clinical trial website goes beyond directory listings with basic descriptive information, such information is considered part of the informed consent process and therefore requires IRB review and approval."[63]

Researchers also use the Internet to observe real-time behavior, for example by entering chat rooms and observing the interactions between the site users, and possibly taking part in the interactions themselves. Wake Forest University researchers recently conducted an IRB-approved study in which members of the

---

157. It should be noted that this type of research is often deemed "exempt" from Subpart A in accordance with 45 C.F.R. § 46.101(b), especially if no names are collected along with the surveys, but exempt research is still human subjects research. 45 C.F.R. § 46.101(b) (2012).

58. *See, e.g.*, Novartis, Clinical Trials and Medical Research Trials Recruiting in the U.S., http://www.novartisclinicaltrials.com/webapp/etrials/home.do (last visited Aug. 9, 2012); Pfizer, Find A Trial, http://www.pfizer.com/research/research_clinical_trials/find_a_trial.jsp (last visited Aug. 9, 2012).

59. CRAIGSLIST, www.craigslist.com (last visited Aug. 9, 2012). Craigslist postings expire after a short period of time and links to specific advertisements become broken; however, Craigslist advertisements for research participants are often found in a city's community section, under "volunteers." *See, e.g.,* http://boston.craigslist.org/vol/.

60. U.S. National Institutes of Health, ClinicalTrials.gov, www.clinicaltrials.gov (last visited Aug. 9, 2012).

61. 45 C.F.R. §46.111.

62. Matthew D. Whalen & Felix A. Khin-Maung-Gyi, *Recruitment of Research Subjects*, *in* IRB MANAGEMENT AND FUNCTION 147, 147 (Elizabeth A. Bankert & Robert J. Amdur 2005).

63. OHRP, Guidance on Institutional Review Board Review of Clinical Trial Websites, http://www.hhs.gov/ohrp/policy/clinicaltrials.html (last visited Aug. 9, 2012) (providing more detailed information about when a posting that advertises for participants on a clinical trial web site requires IRB review).

research team entered chat rooms for men who have sex with men.[64] The study methods included interactions with the chat room visitors and the online administration of surveys to willing respondents.[65] The research team disclosed the purpose of its presence in the chat room and sought consent for the online surveys.[66]

These are just three key ways in which the Internet is used in human subjects research.[67] This article, however, focuses on researchers' use of social networking sites, although some of these uses do not include research with human subjects subject to regulation under 45 C.F.R. § 46.

B. Uses of Social Networking Sites in Research

The purpose of a social networking site is to make individuals available to be contacted and able to contact others for any number of purposes.[68] These sites, which include commonly known sites such as Facebook,[69] MySpace,[70] and Twitter,[71] allow people to post information about themselves, communicate with friends and family, and partake in other activities online, such as playing games.[72] Admissions committees and employers regularly review social networking sites to determine whether a candidate may be suitable for an academic program or job, whether or not the applicant even knows about or has given permission for such a review.[73] The information posted on a social networking site could also prevent a state bar applicant from being admitted to practice law.[74]

---

64. Scott D. Rhodes et al., *A Pilot Intervention Utilizing Internet Chat Rooms to Prevent HIV Risk Behaviors Among Men Who Have Sex with Men*, 125 PUB. HEALTH REPORTS 29 (2010) (Supp. 1).

65. *Id.*

66. *Id.* at 31.

67. Other uses might include re-contacting participants after a study in which they participated months or years before to participate in a new study, or attempting to contact study participants who may be lost to follow-up.

68. Lauren Gelman, *Privacy, Free Speech, and "Blurry-Edged" Social Networks*, 50 B.C. L. REV. 1315, 1326-27 (2009).

69. FACEBOOK, www.facebook.com (last visited Aug. 9, 2012).

70. MYSPACE, www.myspace.com (last visited Aug. 9, 2012).

71. TWITTER, www.twitter.com (last visited Aug. 9, 2012).

72. Powell, *supra* note 23, at 163; Gelman, *supra* note 68, at 1326-27.

73. *See* Jin Pyuo Lee, *Law Schools Check Applicants' Facebook Profiles*, THE DAILY PENNSYLVANIAN, Nov. 13, 2008, *available at* http://thedp.com/index.php/article/2011/11/admissions_offices_check_facebook; Dina Epstein, *Have I Been Googled? Character and Fitness in the Age of Google, Facebook, and YouTube*, 21 Geo. J. Legal Ethics 715, 725 (2008); John Hechinger, *College Applicants Beware, Your Facebook Page is Showing*, WALL ST. J., Sept. 18, 2008, at D1.

74. Kathleen Elliot Vinson, *The Blurred Boundaries of Social Networking in the Legal Field: Just "Face" It*, 41 U. MEM. L. REV. 355, 388-89 (2010).

Social networking site use among the general public is widespread. Facebook reports having over 900 million users,[75] though the number of unique, active users may be lower. In March 2012, Twitter reported having over 140 million users.[76] Researchers are also using social networking sites more frequently as a means of facilitating their research. For example, in addition to using sites like Craigslist to seek study participants,[77] researchers are turning to sites like Facebook to recruit participants for their studies.[78] Researchers also may use social networking sites to locate and contact participants whose progress they were following after an intervention but fell out of touch with the research team.[79]

In addition to these uses of social networking sites, academic researchers are also more frequently engaging in the practice of data mining on social networking sites as a means of collecting data for biomedical, social, and behavioral research. Often, this is their sole method for data collection in a study. For example, in 2011 researchers at the University of Florida mined the Facebook pages of hundreds of medical students and residents at the university to determine how many of the students and residents posted information that might represent a violation of a patient's privacy.[80] They ultimately concluded that there had been "significant and increasing evidence of potential privacy violations"[81] that consisted of mostly medical students, but also residents, posting pictures of themselves with patients who could have been identified from the photographs.[82]

The researchers did not specifically request access to anyone's Facebook page in order to collect the data. Rather, the sole method of data collection was through data mining.[83] The publication reporting the data notes that the researchers accessed the Facebook pages of the medical students and residents via their personal Facebook accounts, viewing the entirety of the publicly

75. Newsroom, FACEBOOK, http://newsroom.fb.com/content/default.aspx?NewsAreaId=22 (last visited Aug. 9, 2012).

76. *Twitter Turns Six,* TWITTER, http://blog.twitter.com/2012/03/twitter-turns-six.html (Mar. 21, 2012, 10:18 EST).

77. *Supra* note 59, discussing Craigstlist advertisements for research participants.

78. *See, e.g.*, SC Clinical Research, FACEBOOK, http://www.facebook.com/pages/SC-Clinical-Research/110492726886 (last visited Aug. 9, 2012); West Coast Clinical Trials, FACEBOOK, http://www.facebook.com/pages/West-Coast-Clinical-Trials/82486549909 (last visited Aug. 9, 2012).

79. *See, e.g.*, Richelle Mychasiuk & Karen Benzies, *Facebook: An Effective Tool for Participant Retention in Longitudinal Research*, CHILD: CARE, HEALTH & DEV., published online Oct. 10, 2011.

80. Lindsay A. Thompson et al., *Protected Health Information on Social Networking Sites: Ethical and Legal Considerations*, 13 J. MED. INTERNET RES. e8 (2011).

81. *Id.*

82. *Id.*

83. *Id.*

available pages, and only viewing the main photograph and home page of those for which privacy settings had been activated.[84]

Researchers at Cornell University reviewed the tweets, messages posted on Twitter, of more than two million people to make conclusions about what time of day people are the happiest.[85] They concluded, perhaps not surprisingly, that people are in the best mood in the morning, and on weekends.[86]

Whether this type of research is considered research with human subjects and therefore requires IRB review hinges on whether or not the research involves "human subjects" as defined in 45 C.F.R. § 46.102(f). According to this provision, a "human subject" is "a living individual about whom an investigator (whether professional or student) conducting research obtains (1)[d]ata through intervention or interaction with the individual, or (2)[i]dentifiable private information."[87] I have previously contended that such research is likely not research with human subjects because it does not involve intervention or interaction with any individual and, more importantly for the purposes of this article, does not involve obtaining private information about a person.[88]

Without any input from HHS on this issue, researchers like those at the University of Florida and Cornell University are left to determine for themselves, or must determine in conjunction with their IRB, whether IRB approval is required in order to conduct their research. If no approval is required, these researchers need not spend the time to put together an IRB application, and the IRB need not spend its limited resources reviewing that which by regulation is not within its purview. Furthermore, if no IRB approval is required, the question of whether researchers may need to obtain informed consent from each individual whose data are mined becomes a contractual issue based on the terms of use of the social networking site, as well as an ethical issue.

IV. PRIVACY, SOCIAL NETWORKING SITES, AND THE LAW

While this article contends that data mining on social networking sites is not human subjects research, it acknowledges that the reason that there is even a question as to whether such research is research with human subjects is because of the lack of clarity about what "private information" really means in today's society, and, in particular, how the law regards reasonable expectations of privacy today. An in-depth analysis of recent case law supports the contention

84. *Id.*

85. Scott A. Golder & Michael W. Macy, *Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Culture*, 333 SCIENCE 1878 (2011).

86. *Id.* at 1878.

87. 45 C.F.R. 46.102(f).

88. Solberg, *supra* note 16, at 321, 324, 328. However, the University of Florida Facebook study was submitted to the institution's IRB and deemed exempt. Thompson, *supra* note 80. The Cornell University researchers did not report whether or not their study was submitted to their IRB. *See* Golder & Macy, *supra* note 85.

that data mining on social networking sites is not human subjects research because individuals do not have a reasonable expectation of privacy with respect to the information they post on these sites.[89]

A. Courts Hold No Reasonable Expectation of Privacy on Social Networking Sites

The federal regulations that govern research with human subjects define private information as "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)."[90] Unfortunately, these regulations do not identify any specific contexts in which individuals may have a reasonable expectation of privacy, and only provide the example of a medical record as information that an individual can reasonably expect will remain private.[91] Furthermore, the OHRP has not published any guidance or provided any other advice that identifies such contexts or information.

However, in the last few years, courts have been asked to decide in civil cases whether individuals have a reasonable expectation of privacy in information that they post on social networking sites.[92] The opinions provide an excellent means of interpreting the way in which "private information" may be contemplated in 45 C.F.R. § 46 in light of the technology currently available and the way the Internet is being used to conduct academic research.[93] Ultimately, courts across the United States are consistently holding that there is no expectation of privacy with respect to the information that people post on social networking sites and have compelled production of this information in a variety of cases.[94]

The approach that these courts have taken – that sharing information on a social networking site, even with only a small group of people, results in an unreasonable expectation of privacy with respect to the information posted – is one that makes sense when viewed in the context of the privileges that the law

89. *See, e.g.*, Jaffee v. Redmond, 518 U.S. 1 (1996); Trammel v. United States, 445 U.S. 40 (1980); Romano v. Steelcase, 907 N.Y.S. 2d 650 (N.Y. Sup. Ct. 2010).

90. 45 C.F.R. § 46.102(f).

91. *Id.*

92. *See, e.g.,* Jaffee v. Redmond, 518 U.S. 1; Trammel v. United States, 445 U.S. 40; Romano v. Steelcase, 907 N.Y.S. 2d 650. Standards and issues relevant to privacy in criminal cases differ significantly from those applicable to the academic research context. However, the standards and issues present in civil cases represent a better mechanism for exploring expectations of privacy online and in academic research. Thus, this article explores civil, rather than criminal, cases.

93. *Id*.

94. *Id*. *See also E-Discovery for Defendants Cheat Sheet*, DRUG & DEVICE LAW BLOG, http://druganddevicelaw.blogspot.com/2011/11/e-discovery-for-defendants-cheat-sheet.html (Nov. 22, 2011, 22:48 EST) (for a detailed list of cases that discuss discovery of information posting on social networking sites).

typically recognizes.[95] A privilege is "an evidentiary rule that gives a witness the option to not disclose the fact asked for, even though it might be relevant; the right to prevent disclosure of certain information in court, especially when the information was originally communicated in a professional or confidential relationship."[96]

It is within the context of privilege that I analyze the meaning of "private information" as used in 45 C.F.R. § 46.102(f). The meaning of private information in 45 C.F.R. § 46 is analogous to the meaning of privilege in the rules of evidence.[97] The Federal Rules of Evidence recognize that there are certain communications between individuals that are privileged.[98] For example, communications between an attorney and client,[99] psychotherapist and patient,[100] and husband and wife[101] may be protected from compelled disclosure. Furthermore, state statutes also may address privileged communications. Tennessee protects communications between, among others, attorney and client,[102] social worker and patient,[103] psychiatrist and patient,[104] husband and wife,[105] and accountant and client.[106] Any information communicated to individuals under circumstances other than those outlined in these statutes is not protected by the statutory privilege and is subject to disclosure.[107]

A closer look at these recent cases provides greater insight into how courts view the possibility that privileged information can be posted online, and how these views can help researchers and IRBs determine whether research involves the collection of private information as contemplated in 45 C.F.R. § 46. These cases furthermore provide insight into why information posted on social networking sites is not "information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place," as private information is also defined in 45 C.F.R. § 46.102(f).

In *Romano v. Steelcase*, the defendant in a personal injury case sought access to the plaintiff's MySpace and Facebook accounts in an effort to dispute the plaintiff's claims of injury and loss of enjoyment of life.[108] The plaintiff

95. *See infra* notes 99-106.
96. BLACK'S LAW DICTIONARY 1215 (7th ed. 1999).
97. *Compare* 45 C.F.R. § 46.102(f) ("Private information includes . . . 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") *with* BLACK'S LAW DICTIONARY, *supra* note 96, at 1215.
98. FED. R. EVID. 501.
99. FED. R. EVID. 502.
100. Jaffee v. Redmond, 518 U.S. 1 (1996).
101. Trammel v. U.S., 445 U.S. 40 (1980).
102. TENN. CODE ANN. §23-3-105 (2011).
103. TENN. CODE ANN. §63-23-109 (2011).
104. TENN. CODE ANN. §24-1-207 (2011).
105. TENN. CODE ANN. §24-1-201 (2011).
106. TENN. CODE ANN. §62-1-116 (2011).
107. FED. R. EVID. 501.
108. Romano v. Steelcase, 907 N.Y.S. 2d 650, 651 (N.Y. Sup. Ct. 2010).

asserted her right to privacy regarding her social networking site postings, and thus argued that she should not be required to produce the information from her accounts (including old, deleted postings).[109] The court held that, "notwithstanding her privacy settings,"[110] the plaintiff consented to sharing the information that she posted on these sites, as the primary purpose of social networking sites is to share information with others.[111] The court would therefore not allow her to claim an expectation of privacy with respect to information that was intentionally shared with non-privileged individuals.[112] It furthermore cited other civil cases in which courts have similarly held that people do not have reasonable expectations of privacy with respect to social networking site postings that they have shared with others,[113] and, therefore, the court concluded that no expectation of privacy existed in that case.

One of the cases the court cited was *Beye v. Horizon Blue Cross Blue Shield of New Jersey*, a case alleging breach of contract by an insurance company, in which the court ordered the plaintiffs to produce Facebook and MySpace postings discussing their eating disorders because these postings were shared, rather than kept private (like diary entries).[114] The court ordered this discovery despite the fact that the postings addressed sensitive information about the plaintiffs.[115] Thus, one can compare *Beye* to a situation where a researcher intends to mine sensitive data from a site like Facebook or MySpace, perhaps about health issues such as eating disorders, and conclude that despite the sensitive nature of the information posted, it is not private information as defined in 45 C.F.R § 46.

A New York court cited *Romano* in another personal injury case, holding that Facebook postings, if relevant, are still subject to discovery even though privacy settings prevented the general public from accessing the postings.[116] Likewise, in Pennsylvania, the court in *Largent v. Reed* held that "there is no reasonable expectation of privacy in material posted on Facebook."[117] This was also a personal injury case, with the plaintiffs in *Largent* alleging physical injury and pain and suffering as the result of an automobile accident.[118] The defendant sought access to one plaintiff's Facebook page because he believed it would reveal photos of the plaintiff at the gym after the accident, thus showing her

109. *Id.* at 651, 655.
110. *Id.* at 657.
111. *Id.*
112. *Id*.
113. *Id.* at 656 (quoting Beye v. Horizon Blue Cross Blue Shield of New Jersey, No. 06-5337, 2007 U.S. Dist. LEXIS 100915 (D.N.J. December 14, 2007) ("The privacy concerns are far less where the beneficiary herself chose to disclose the information.").
114. Beye, 2007 U.S. Dist. LEXIS 100915, at *9 n3.
115. *Id.*
116. Patterson v. Turner Const. Co., 931 N.Y.S. 2d 311, 312 (N.Y. Sup. Ct. 2011).
117. Largent v. Reed, No. 2009-1823, at *9 (Pa. C.P. Franklin Co. Nov. 7, 2011).
118. *Id.* at *2.

claims of injury to be without merit.[119] The court held that "no general privacy privilege protects [plaintiff's] Facebook material from discovery"[120] and that she has no reasonable expectation of privacy protecting her Facebook page because "[a]lmost all information on Facebook is shared with third parties, and there is no reasonable privacy expectation in such information."[121]

In 2010, in *McMillen v. Hummingbird Speedway*, another Pennsylvania court also refused to recognize any kind of social networking site privilege.[122] The case was a personal injury case arising from a stock car accident, in which the defendant sought during discovery to obtain log-in information for the plaintiff's Facebook account to examine whether the extent of the plaintiff's injuries were as serious as he claimed.[123] The court opined that, particularly in light of the privacy policies and terms of use that sites like Facebook and MySpace have,[124] "while it is conceivable that a person could use [social networking sites] as forums to divulge and seek advice on personal and private matters, it would be unrealistic to expect that such disclosures would be considered confidential."[125] This court concluded that because "[t]he law does not even protect otherwise privileged communications made in the presence of third parties,"[126] communications made between non-privileged individuals are "wholly incommensurate with a claim of confidentiality."[127]

A third Pennsylvania court following *McMillen* also held that there is no reasonable expectation of privacy in social networking site postings.[128] The plaintiff in *Zimmerman v. Weis Markets* was an employee of the defendant company when he injured himself.[129] Although he alleged serious injury, the defendant saw information and photographs posted on public areas of the plaintiff's Facebook page that indicated that his injuries were not so serious, and accordingly sought access to additional material not shared with the general public.[130] The court held that because no privilege exists that protects the privacy of the information, and the purpose of a social networking site is to share information with others, no reasonable expectation of privacy exists with respect to the information that is posted.[131] Furthermore, those with whom the

119. *Id.* at *5-6.
120. *Id.* at *9.
121. *Id.*
122. McMillen v. Hummingbird Speedway, Inc., No. 113-2010 CD, 2010 Pa. Dist. & Cnty. Dec. LEXIS 270 (Pa. C.P. Jefferson Co. Sept. 9, 2010).
123. *Id.* at *1, 6
124. *Id.* at *3-4.
125. *Id.* at *3.
126. *Id.* at *5.
127. *Id.*
128. Zimmerman v. Weis Markets, Inc, No. CV-09-1535, 2011 Pa. Dist. & Cnty. Dec. LEXIS 187, at *9 (Pa. C.P. Northumberland Co. May 19, 2011).
129. *Id.* at *1.
130. *Id.* at *2.
131. *Id.* at *4, 6.

information is shared are not restricted in terms of how they may in turn share the posted information.[132]

These opinions support the notion that "private information," as the term is used in 45 C.F.R. § 46.102(f), does not include information posted by individuals on social networking site pages, and that research involving data mining on social networking sites therefore need not be subject to IRB review because the researchers are not obtaining identifiable, *private* information about individuals.[133] However, these cases are not the only ones that support the contention that mining data on social networking sites is not human subjects research.

In a child custody case, an Ohio court refused to recognize a woman's right to privacy regarding her MySpace and other blog postings about illegal drug use when it considered whether the trial court erred in considering this evidence when it awarded custody to her daughter's father.[134] It should be noted that this mother told the court that any member of the public could read these postings, and the court accordingly concluded that she could not assert a right to privacy with respect to them.[135]

Of course, if such blog postings were to be used for research purposes – rather than as evidence in a child custody case, which is what the court was hearing – there are ethical considerations that the researcher should need to take into account at the time of publication. For example, the researcher should consider whether publication of information in a particular way, *e.g.,* by directly quoting the blog post, could lead to the identification of the blog writer. This is especially true when blog posts – such as the ones in this case – discuss engaging in illegal acts.[136] Furthermore, the same analysis holds if the information does not reveal illegal behavior but instead pertains to someone's health condition.[137] If a researcher has information about someone and the person could be identified based on the way the information is presented in the publication, even if no individual names are ultimately included in that publication, a researcher should consider the potential consequences that may result of the identification. However, the ethical issues associated with the possible identification of a data source do not necessarily lead to the conclusion that the research is human subjects research and requires IRB review.

In *Moreno v. Hanford Sentinel*, a California court similarly held that people have no reasonable expectation of privacy with respect to information posted

---

132. *Id.* at *8.

133. *See supra* text accompanying note 88.

134. Dexter v. Dexter, No. 2006-P-0051, 2007 WL 1532084, at *6-7 (Ohio App. May 25, 2007).

135. *Id.* at *6, n4.

136. *See id.*

137. *See* Beye v. Horizon Blue Cross Blue Shield of N.J., No. 06-5337, 2007 U.S. Dist. LEXIS 100915 (D.N.J. December 12, 2007).

publicly on the Internet.[138] A woman wrote an article about the town in which she grew up. In the article, she included various unfavorable statements about the town and the people who lived there.[139] She then posted it on her MySpace page, and it was available for public viewing, *i.e.,* without any specific access to her page or passwords required.[140] A school principal from the town sent it to the local newspaper before the woman had removed it from her page, and the newspaper later published it, citing her as the author.[141] The woman sued the principal and the newspaper for invasion of privacy and the intentional infliction of emotional distress.[142] The court held that by posting the article on her MySpace page, the plaintiff "made her article available to any person with a computer and thus opened it to the public eye. Under these circumstances, no reasonable person would have had an expectation of privacy regarding the published material."[143]

The court further opined that "the fact that [the plaintiff] expected a limited audience does not change [this] analysis."[144] Thus, blog postings that can be read by anyone with Internet access should not be considered private, even if the writer expects that only a small number of individuals will view the blog.[145] Furthermore, the court stated that the fact that the posting was online for less than a week before it was removed was irrelevant and held that the posting "was not so obscure or transient that it was not accessed by others."[146] Thus, if a researcher collected information from blog posts that had only been posted a short amount of time and were removed fairly soon after being posted, the author may have a difficult time justifying a reasonable expectation of privacy simply because they were posted only for a brief period of time.[147]

These decisions are helpful in evaluating the issue of whether researchers who mine and analyze information posted on social networking sites need to submit their research to their IRBs for review and approval. The cases can be applied to analyze, for example, the question that Internet research ethics scholar Elizabeth Buchanan posed in her presentation about Internet research in the Summer of 2010 to the Secretary's Advisory Committee on Human Research Protections.[148] In this presentation, she asked whether a researcher who analyzes

---

138. Moreno v. Hanford Sentinel, Inc., 172 Cal. Ct. App. 4th 1125, 1130 (2009) (cited in Gelman, *supra* note 68, at 1340-41).
139. *Id.* at 1128.
140. *Id.* at 1128, 1130.
141. *Id.* at 1128.
142. *Id.* at 1129.
143. *Id.* at 1130.
144. *Id.*
145. Gelman, *supra* note 68, at 1341.
146. *Moreno*, 172 Cal. Ct. App. 4th at 1130.
147. *See id*.
148. Elizabeth Buchanan, Presentation to the Secretary's Advisory Committee on Human Research Protections, *Internet Research Ethics and IRBs*, July 21, 2010, available at http://www.hhs.gov/ohrp/sachrp/mtgings/mtg07-10/present.html.

blog postings must submit the study for IRB review or risk the inability to publish the work.[149] The analysis that can be used to answer this question is no different from the analysis that can be used to determine whether individuals who post on Facebook or Twitter have a reasonable expectation of privacy, and the essential part of this analysis again is whether the information contained in the blog posts to be analyzed is private.[150] It would therefore appear that Professor Buchanan's question can be answered using the same analysis that the courts have used to answer the question of whether information posted on social networking sites is private.

B. Reasonable Expectations of Privacy: Counter-Arguments

The fact that courts are recognizing that sharing information with even a limited number of third parties reduces or eliminates any reasonable expectation of privacy that an individual can have with respect to information they post on a social networking site means that it is reasonable to interpret "private information" as described in 45 C.F.R. § 46 to exclude information posted on a social networking site. However, the argument has been posited that people who post on social networking sites may have some reasonable expectation of privacy with respect to the posted information if they limit the number of people who may access the information, *e.g.,* by adjusting the privacy settings on their accounts.[151]

Internet scholar Michael Zimmer, for example, addressed this argument in an article where he criticizes the way in which a data set that Harvard University researchers created based on information posted on the Facebook pages of the students in Harvard's class of 2009 was published.[152] These researchers, who created the data set to make it available to other researchers, claimed that the data had been de-identified, and Professor Zimmer explains how easily the individuals whose information had been mined and recorded could be re-identified.[153]

Professor Zimmer was furthermore concerned about the way in which data was collected from the individual Facebook pages to compile this data set due to the expectations of privacy that those whose information was recorded may have had with respect to their Facebook pages.[154] The Harvard researchers were able

---

149. *Id.*

150. In accordance with federal regulations, the information must also be identifiable, but for purposes of this analysis, I am presuming that the information is identifiable and therefore the only remaining issue is whether it is private. The ability or inability to identify an individual who posts information online is outside of the scope of this article.

151. *See, e.g.*, Michael Zimmer, *"But the data is already public": On the Ethics of Research in Facebook*, 12 ETHICS INFO. TECH. 313, 318 (2010) (discussing Kevin Lewis et al., *Tastes, Ties, and Time: A New Social Network Dataset Using Facebook.com*, 30 SOCIAL NETWORKS 330 (2008)).

152. *See generally id.*

153. *Id.* at 316-17.

154. *Id.* at 318.

to mine the Facebook data with the help of research assistants who otherwise had access to the individual Facebook pages, likely because they were all part of the same network(s).[155] Professor Zimmer argues that because some individuals may have adjusted their privacy settings to only allow certain individuals to access their Facebook pages, the fact that the research assistants were part of a group that was given limited access does not mean that these individuals intended for their information to then be used by the research assistants for research purposes.[156] He further contends that such use "puts the privacy of those subjects at risk."[157] Other scholars have similarly contended that "a regime that fails to recognize that not everything posted on the Internet is meant for every user of the Internet, fails to understand the technology and its potential."[158]

One court may agree. Although *Moreno* held that there is no expectation of privacy in information posted on the Internet that is available to the general public,[159] the court nevertheless conceded that, "information disclosed to a few people may remain private."[160] This statement supports arguments like Professor Zimmer's that those who limit access to their Facebook page to a small group of Friends, for example, have a reasonable expectation of privacy with respect to that information because they activated their privacy settings to restrict access to only a select group of people (rather than making their page completely public).[161] However, the *Moreno* court appears to be the exception with respect to that analysis, as both *Largent* and *Romano* held that disclosure of information to any third parties – even if only to a small group – on a social networking site is enough to destroy any reasonable expectation of privacy with respect to that information.[162]

The Stored Communications Act[163] has also been argued as a reason that individuals have a reasonable expectation of privacy in postings on social networking sites. The Act prohibits companies like Facebook and MySpace from being compelled to produce information in civil cases such as private email and instant messages that were exchanged between site users through the sites, as both Facebook and MySpace provide electronic communication services and remote computing services.[164] Such a prohibition has led parties to argue that

---

155. *Id.*

156. *Id.* ("While the information was indeed available to the R[esearch] A[ssistant], it might have been accessible only due to the fact that the RA was within the same 'network' as the subject, and that a privacy setting was explicitly set with the intent to keep that data within the boundaries of that network.")

157. *Id.*

158. Gelman, *supra* note 68, at 1344.

159. Moreno v. Hanford Sentinel, Inc., 172 Cal. Ct. App. 4th 1125, 1130 (2009).

160. *Id.* (quoting M.G. v. Time Warner Inc., 89 Cal. App. 4th 623, 632 (2001)).

161. *See* Zimmer, *supra* note 151.

162. Largent v. Reed, No. 2009-1823, at *9 (Pa. C.P. Franklin Co. Nov. 7, 2011); Romano v. Steelcase, 907 N.Y.S. 2d 650, 656-57 (Sup. Ct. 2010).

163. Stored Communications Act, 18 U.S.C. §§ 2701 to 2712 (2012).

164. *See* Crispin v. Christian Audigier, Inc., 717 F. Supp. 2d 965, 969 (C.D. Cal. 2010).

they have a reasonable expectation of privacy in postings they have made on Facebook,[165] as well as a reasonable expectation of privacy with respect to private messages exchanged between Facebook or MySpace users,[166] since the companies are prohibited by federal law from disclosing the exchanges. However, it should be emphasized that the Stored Communications Act only applies to entities that are providers of electronic communication services or remote computing services.[167]

Interestingly, the court in *Crispin v. Christian Audigier, Inc.* held that there was not enough evidence for it to decide whether or not the Facebook wall messages and MySpace comments could be compelled,[168] but any decision by the court that these comments could be compelled may affect the analysis of whether data mining on social networking sites by academic researchers is human subjects research.

I contend that sharing information with even a select group of people on a social networking site reduces the expectation of privacy that one can have with respect to that information, and as a result, researchers who mine data on social networking sites are not conducting research with human subjects. However, my disagreement does not mean that that the ethical conduct of the research is unimportant.[169] It is important that basic principles of research ethics guide researchers so that they respect the privacy of the individuals whose data they are mining (even if what they are mining is not "private information").[170] However, whether or not there are ethical concerns is an issue separate and apart from whether IRB review is required.

V. A NEW KIND OF SOCIAL NETWORKING SITE

Although this article has discussed several publications whose authors analyzed data collected from social networking site postings,[171] not all social networking sites fit the mold of Facebook, MySpace, and Twitter. Thus, the changing nature of social networking sites provides a persuasive reason for HHS to carefully consider whether data mining on social networking sites is human subjects research. The conclusion that data mining on sites like these to conduct academic research is based in part on the premise that the purpose of a social networking site is to share information with others,[172] but also on the premise that social networking sites like these allow almost anyone to register and create

165. *Largent*, No. 2009-1823, at *10. *See Crispin*, *supra* note 122, at 991.
166. *See Crispin*, 717 F. Supp. 2d at 991.
167. *See* Stored Communications Act § 2702.
168. *Crispin*, 717 F. Supp. 2d at 991.
169. *See supra* text accompanying note 136-37 (describing the importance of conducting research ethically even though it may not be human subjects research).
170. *See* Zimmer, *supra* note 151, at 324 ("[I]t is our responsibility [] to ensure our research methods and processes remain rooted in long-standing ethical practices.")
171. *See, e.g.,* Golder & Macy, *supra* note 85; Thompson, *supra* note 80.
172. Gelman, *supra* note 68, at 1326.

an account where they can post information for others to view.[173] There is, however, a new social networking site that is increasing in popularity that forces a more nuanced approach to determining whether data mining on social networking sites is human subjects research subject to regulation under 45 C.F.R. §46.

Founded in 2004, PatientsLikeMe[174] provides users with more than just a general social networking opportunity, and is indeed more restrictive regarding who may become a user than Facebook, MySpace, or Twitter, although it has been described as appearing "[a]t first glance. . . like just any other online community, a kind of MySpace for the afflicted."[175] Now with more than 150,000 users,[176] PatientsLikeMe was designed to provide patients with the tools to create a profile where they can record information about their health, analyze their own health data, and then compare it to others.[177] Although data mining on this type of website will not serve as a complete substitute for traditional Phase I-IV clinical trials, academic researchers can gather important data from patients who post information about themselves and their health care online.[178]

Given the public's widespread use of the Internet to learn more about their health and health care,[179] it is unsurprising that a health care-specific social networking site has been developed, particularly given the popularity of sites like WebMD that people turn to regularly (for better or for worse) for information about their health.[180] The abundance of articles that the PatientsLikeMe employees/researchers have published to date based on data collected from participants on the site (whether through data mining or surveys) clearly

---

173. Facebook, for example, allows anyone over the age of 13 to register for a Facebook account. *Statement of Rights and Responsibilities*, FACEBOOK, http://www.facebook.com/legal/terms?ref=pf (last visited Aug. 9, 2012). Twitter requires that its users be able to "form a binding contract with Twitter" and prohibits use of the site by anyone who by law has been banned from using it. *Terms of Service*, TWITTER, http://twitter.com/tos (last visited Aug. 9, 2012). In the United States, this effectively means that almost everyone over the age of 18 can register for a Twitter account.

174. PATIENTSLIKEME, www.patientslikeme.com (last visited Aug. 9, 2012).

175. Thomas Goetz, *Practicing Patients*, N.Y. TIMES, Mar. 23, 2008, at MM32.

176. PATIENTSLIKEME, *supra* note 174.

177. Gerald C. Kane et al., *Community Relations 2.0*, Harv. Bus. Rev., Nov. 2009, at 45, 46("Patients volunteer details about their diseases and the treatments they've pursued – including those not prescribed by their doctors. Charts and progress curves on the website help people to visualize their own complex treatment histories, allow comparisons among peer groups, and prompt members to provide feedback and advice on one another's progress."); *id.*

178. Amy Dockser Marcus, *ALS Study Shows Social Media's Value as Research Tool*, WALL ST. J., Apr. 25, 2011, at A5.

179. Kane, *supra* note 177, at 46 ("A report from Manhattan Research suggests that more than 60 million Americans are consumers of 'health 2.0' resources. They read or contribute to blogs, wikis, social networks, and other peer-produced efforts, using Google as the de facto starting point.").

180. The *New York Times* has referred to WebMD as an "enormous and powerful site" that is also a "hypochondria time suck." Virginia Heffernan, *A Prescription for Fear*, N.Y. TIMES, Feb. 4, 2011, at MM14.

indicates that this is a resource ripe for mining by other researchers.[181] In 2011, for example, the researchers mined data from the site about patients who reported taking two particular drugs for an off-label purpose (*i.e.,* a use not approved by the FDA), and published an article with their conclusions about the frequency and types of off-label uses and the reported side effects of these medications.[182]

What is unclear, however, is whether mining sites like PatientsLikeMe should be treated like mining Facebook (and therefore is not human subjects research) or if data mining on a site like PatientsLikeMe requires that a different conclusion be reached. It is the unique nature of PatientsLikeMe that forces this question, which depends on whether the expectations of privacy of PatientsLikeMe users differ from – or perhaps more important, should differ from – the expectations of privacy of Facebook or MySpace users.

PatientsLikeMe, unlike Facebook or MySpace, requires users to certify that they are members of one or more groups, such as a person with a particular disease, or a caregiver of a person with a particular disease, in order to register.[183] A health care professional, including one who is a "health researcher," is also permitted to register on the site.[184] However, the site's User Agreement lacks provisions addressing what uses a health researcher could make of the information posted on site users' profiles.[185] The User Agreement only prohibits the use of the site for "commercial purposes," prohibiting "organizations, companies, and/or businesses," as well as "representatives from life sciences and insurance companies" from becoming members, using the site, or creating profiles.[186] No definition of "commercial purposes" is provided, but a biomedical researcher employed at a university who mines data on PatientsLikeMe solely for the purpose of preparing an academic publication is likely not using the site for commercial purposes, *i.e.,* to make money. Furthermore, on the site's R&D policy, which is simply a list of FAQs about conducting research using the site, PatientsLikeMe separates questions about

181. A February 2012 search on PubMed for articles that included the term "PatientsLikeMe" yielded nineteen articles, fifteen of which included at least one author who is an employee of PatientsLikeMe. Of the four remaining articles, only one mined data on PatientsLikeMe to assess the medical terminology used, with a goal of augmenting an online health vocabulary application. *See* Kristina Doing-Harris & Qing *Zeng-Treitler,* Computer-assisted update of a consumer health vocabulary through mining of social network data, 13 J. MED. INTERNET RES. e13 (2011).

182. Jeana Frost et al., *Patient-reported Outcomes as a Source of Evidence in Off-Label Prescribing: Analysis of Data from PatientsLikeMe*, 13 J. MED. INTERNET RES. e6 (2011). The authors of the article are employees of PatientsLikeMe, and not from an academic institution.

183. *User Agreement*, PATIENTSLIKEME, http://www.patientslikeme.com/about/user_agreement (last visited Aug. 9, 2012).

184. *Id.*

185. *Id.*

186. *Id.*

"commercial research" from questions about "non-commercial/academic research."[187]

Thus, one can conclude that the User Agreement does not anticipate that academic researchers are using the site for commercial purposes, and that the use of PatientsLikeMe for academic research purposes is therefore permissible without explicit permission from the site.[188] If explicit permission from PatientsLikeMe is not required for academic researchers to mine the site for research purposes, PatientsLikeMe users arguably have a lower expectation of privacy with respect to the information they post on the site than they would if explicit permission were required. This is because if explicit permission were required, the pool of individuals who could mine their information would likely be smaller than the pool who may mine their information when such permission is not required.

The research published by PatientsLikeMe employees also provides some insight into the expectations of the site users with respect to research. In an article about multiple sclerosis patients and their barriers to adhering to their medication regimens which reported the results of surveys completed by registered PatientsLikeMe site users, the authors state that "[m]embers of PatientsLikeMe join the site with the expectation that they will be participating in research."[189] The authors do not take this statement one step further to clarify whether this means that individuals who join PatientsLikeMe expect that they will be asked to participate in surveys, or if individuals who join expect that the information they post will be used for research purposes, or both. However, if PatientsLikeMe users are joining with the expectation that, in general, they will be research participants by virtue of their membership on the site, they again

187. PatientsLikeMe, Frequently Asked Questions about research at PatientsLikeMe, http://patientslikeme_research.s3.amazonaws.com/Research%20Policy%20May2009.pdf (last visited Aug. 9, 2012).

188. Unfortunately, PatientsLikeMe does not state its policy on what types of approval are required from the site to conduct independent academic research; instead, it only comments on academic research done in collaboration with PatientsLikeMe. *Id.* By way of comparison, Facebook's Statement of Rights and Responsibilities provides that "If you collect information from users, you will: obtain their consent, make it clear you (and not Facebook) are the one collecting their information, and post a privacy policy explaining what information you collect and how you will use it." Statement of Rights and Responsibilities, FACEBOOK, http://www.facebook.com/legal/terms (last visited Aug. 9, 2012). However, much like on PatientsLikeMe, an ambiguity likely exists for individual users about the meaning of this provision. Facebook does not define the term "collect" and there are certainly a number of uses that are made of Facebook for which consent of individuals is not sought but their information may be considered "collected." *See supra* notes 73-74. Thus, it is unlikely that a Facebook user could claim a reasonable expectation of privacy regarding information posted on her page because of this provision in the Statement of Rights and Responsibilities. Should Facebook amend this provision in the future, further investigation about expectations of privacy of Facebook users may be warranted.

189. Paul Wicks et al., *Use of an Online Community to Develop Patient-Reported Outcome Instruments: The Multiple Sclerosis Treatment Adherence Questionnaire (MS-TAQ)*, 13 J. MED. INTERNET RES. e12 (2011).

should have reduced expectations of privacy with respect to the information they post.

It would seem that notions of reasonable expectations of privacy change may when people post information specifically about their health online. Laws like the Health Insurance Portability and Accountability Act (HIPAA) exist to prevent the unauthorized disclosure of an individual's protected health information by third parties such as doctors, nurses, and insurance companies (what are called, under HIPAA, "covered entities").[190] However, there are no HIPAA implications with respect to posting information on PatientsLikeMe because the site is not a covered entity.[191] Additionally, "health information stored by a patient on an online health profile – or even a personal filing cabinet – has no claim to privacy."[192] Therefore, PatientsLikeMe users have an ever further reduced expectation of privacy in the information they post.

It is therefore difficult to say with certainty that data mining on a site like PatientsLikeMe is human subjects research under 45 C.F.R. § 46 because placing restrictions on who can join a site is indicative of a more reasonable expectation of privacy that individuals. However, these restrictions do not eliminate health researchers from joining the site and mining the data on it for research purposes. Furthermore, PatientsLikeMe's statement that people who join the site expect to be research participants could be viewed as a warning to potential and current site users that use of the site comes with the possible "risk" of being a research participant.[193] Coupled with the argument that individuals cannot expect that health information they disclose to someone other than a covered entity will remain private, a court could determine that despite the stricter registration requirements on a site like PatientsLikeMe, if the User Agreement specifically allows researchers to join the site and consequently have access to all users' profiles, and if site users join knowing they will be research participants, information posted on the site would not be considered private information. By extension, under 45 C.F.R. § 46, it is unlikely that mining that information would be human subjects research.

---

190. Health Insurance Portability and Accountability Act, 42 U.S.C. §1320d (2012).

191. If PatientsLikeMe was a covered entity, it would be legally obligated to protect the health information shared by its users, but it is not a covered entity because it is not a health care provider. James Grimmelmann, *PatientsLikeMe: A Study in Online Community Issues*, PRAWFSBLAWG, http://prawfsblawg.blogs.com/prawfsblawg/2008/03/the-latest-new.html (Mar. 24, 2008, 9:30 EST).

192. Patricia Sanchez Abril & Anita Cava, *Health Privacy in a Techno-Social World: A Cyber-Patient's Bill of Rights*, 6 NW. J. TECH. & INTELL. PROP. 244, 262 (2008).

193. PatientsLikeMe, Frequently Asked Questions about research at PatientsLikeMe, http://patientslikeme_research.s3.amazonaws.com/Research%20Policy%20May2009.pdf (last visited Aug. 9, 2012).

VI. SUGGESTED LANGUAGE FOR GUIDANCE OR REGULATORY PROVISIONS

HHS understands that the nature of human subjects research has changed since the 1970s, and even since the Common Rule was codified in 1991.[194] The agency furthermore acknowledges that the Internet in particular has played a major role in the way such research is conducted now as compared to more than twenty years ago.[195] In July 2010, a panel was convened at the meeting of the Secretary's Advisory Committee on Human Research Protections (SACHRP) to inform the SACHRP about how the Internet is used in human subjects research, and how the Internet affects our understanding of the way the regulations now must be applied to proposed research.[196]

One year later, in July 2011, HHS published an advance notice of proposed rulemaking (ANPRM) in the Federal Register that sought comments "on how to better protect human subjects who are involved in research, while facilitating valuable research and reducing burden, delay, and ambiguity for investigators."[197] This was the first time since 1988 that HHS publicly requested comments about significant revisions to 45 C.F.R. § 46.[198] HHS again recognized the changes that the Internet has caused in the way human subjects research is conducted, admitting that

> [a]lthough the regulations have been amended over the years, they have not kept pace with the evolving human research enterprise, the proliferation of multi-site clinical trials and observational studies, the expansion of health services research, research in the social and behavioral sciences, and research involving databases, *the Internet*, and biological specimen repositories, and the use of advanced technologies, such as genomics.[199]

---

194. The 1991 final rule that represents the most recent revision of the Common Rule was published following a 1988 notice of proposed rulemaking that proposed a set of regulations that would apply to research funded by agencies other than just HHS (and would therefore be "common" to these federal agencies, thus resulting in the term "the Common Rule."). Federal Policy for the Protection of Human Subjects, 53 Fed. Reg. 45660, 45661 (Nov. 10, 1988) (to be codified at 45 C.F.R. pt. 46).

195. Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 Fed. Reg. 44512, 44513 (July 26, 2011).

196. Buchanan, *supra* note 148. *See* text accompanying note 148 *supra* (discussing Professor Buchanan's presentation to the Secretary's Advisory Committee on Human Research Protections in July 2010).

197. Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 Fed. Reg. 44512 (July 26, 2011).

198. Although notices of proposed rulemaking have been published since 1988, none have sought to make the major changes to 45 C.F.R. § 46 that the 1988 notice and 2011 advance notice proposed.

199. Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 Fed. Reg. at 44512 (emphasis added).

Furthermore, HHS acknowledges that "[t]he advent of sophisticated computer software programs, the Internet, and mobile technology have created new areas of research activity, particularly within the social and behavioral sciences, exponentially increasing the amount of information available to researchers, while providing the means to access and analyze that information."[200] It will likely be quite some time before HHS publishes a notice of proposed rulemaking in response to the comments solicited from the ANPRM, but it is possible, based on HHS's acknowledgements in the ANPRM of the effect the Internet has had on human subjects research, that privacy issues as they relate to the Internet will be addressed.[201] To that end, I am proposing some language that represents what I believe HHS should incorporate into any revisions of 45 C.F.R. § 46, or that the OHRP could incorporate into guidance for researchers and IRBs. This language reflects the changes that the Internet has brought about in human subjects research.

With respect to the use of the Internet to distribute surveys, I propose the following new language:

> The distribution of a survey via the Internet may be exempt under 45 C.F.R. § 46.101(b)(2) or 45 C.F.R. § 46.101(b)(3). However, for all non-exempt research, the distribution of a survey via the Internet requires informed consent obtained in accordance with 45 C.F.R. § 46.116(a)-(b) unless the IRB determines that the research satisfies the criteria for a waiver or alteration of informed consent under 45 C.F.R. § 46.116(c) or 45 C.F.R. § 46.116(d).

With respect to the use of the Internet to recruit research participants, HHS could simply add provisions that incorporate the guidance that the OHRP published in 2005 about advertisements on clinical trial web sites.[202] However, this guidance should be expanded to apply not just to web advertisements for clinical trials, but for web (and possibly even paper) advertisements for any type of research with human subjects.[203]

The OHRP has already stated that such advertisements, if they include more than basic information about a trial, are part of the informed consent process and

---

200. *Id.* at 44513.

201. One scholar is not so hopeful, however. Professor Alexander Halavais, the President of the Association of Internet Researchers, believes that "[t]he proposed changes should reduce many of the burdens currently imposed on researching open discourse on the web – such as studying blog posts or tweets. It is, however, unlikely to make standards of privacy clearer. . . ." Alexander Halavais, *Open Up Online Research*, 480 NATURE 174, 175 (2011).

202. OHRP, Guidance on Institutional Review Board Review of Clinical Trial Websites, *supra* note 63.

203. The guidance was published because the HHS Office of the Inspector General recommended that the OHRP do so, and the recommendation was made specifically with reference to clinical trial web sites. However, the Internet is used to recruit for more than just clinical trials. *See, e.g.,* Mary Ann Chiasson et al., *HIV Behavioral Research Online*, 83 J. URBAN HEALTH 73 (2006); Robert Kraut et al., *Psychological Research Online*, 59 AMER. PSYCHOL. 105, 106 (2004) (identifying web sites where advertisements for research participants may be located).

therefore must be reviewed by the IRB.[204] Furthermore, the OHRP identifies what it considers "basic information."[205] This guidance provides excellent assistance to IRBs and researchers about what needs to be submitted to the IRB, and what the IRB should consider when it reviews the advertisement.[206] The descriptive information included in the guidance about what IRBs should consider when reviewing these web advertisements – such as the way in which incentives are described – is most appropriately left to guidance. However, HHS's recognition in the ANPRM of the widespread use of the Internet for research purposes should be taken one step further into the drafting of a new provision that reads as follows, nearly identically to what is currently written in the OHRP guidance:

> When information posted on a web site to recruit participants in human subjects research goes beyond directory listings with basic descriptive information, such information is considered part of the informed consent process and therefore requires IRB review and approval. Basic descriptive information includes: 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. Information exceeding such basic listing information includes descriptions of study risks and potential benefits, or solicitation of identifiable information. All documents that will be distributed to participants in order to obtain their informed consent must be reviewed by the IRB.[207]

With respect to data mining online, OHRP guidance or HHS regulations should include language similar to the following:

The following activities on the Internet are not human subjects research:

1. Searching for and recording information about individuals that is posted on any website that does not require a password in order to view the site.
2. Searching for and recording information about individuals that is posted on any website if the individual posts the information with the specific designation that it be viewable to the public (for example, a specifically designated public Facebook page).
3. Searching for and recording information about individuals that is posted on any website for which a password is required to enter the web site and/or is not otherwise viewable by the general public if the web site does not require any certification by the researcher of his purpose for using the web site.

---

204. OHRP, Guidance on Institutional Review Board Review of Clinical Trial Websites, *supra* note 63.

205. *Id.* ("Basic descriptive information includes: study title; purpose of the study; protocol summary; basic eligibility criteria; site study location(s), and; how to contact the study site for further information.")

206. *Id.*

207. *See id.* Note again that this language that I am proposing is nearly identical to the language in the current guidance.

4. The use of spyware, software, or other covert technological means such as "hacking" to access information online is not permitted without the permission of the IRB.

VII. CONCLUSION

The Internet is used in many ways to facilitate research, and in particular to facilitate research with human subjects. Based on the current definition of a human subject in 45 C.F.R. § 46, it does not appear that data mining on social networking sites is human subjects research because it involves neither intervention nor interaction with individuals about whom data are obtained. In light of current reasonable expectations of privacy regarding uses of social media, it furthermore does not involve the collection of identifiable, private information about individuals. However, even if data mining on social networking sites is not human subjects research, it is nevertheless important that research facilitated by social media be conducted ethically.[208] At the same time, though, institutions and IRBs must be careful not to conflate the issues of ethical research and research that is subject to regulation under 45 C.F.R. § 46. Simply because research involves the analysis of sensitive personal information does not mean that the research necessarily involves the analysis of identifiable private information.

To reduce the confusion of researchers and IRBs, HHS, as part of its reconsideration of 45 C.F.R. § 46 in its current form, should consider revising the regulations or supplementing them with guidance that addresses the meaning of private information in light of the technological age in which we live. Furthermore, HHS should promulgate regulations that address Internet research, and specifically data mining on sites like social networking sites, or publish guidance to assist researchers and IRBs with understanding when such research is human subjects research and requires IRB review (or at a minimum, a designation of an exemption).

208. Zimmer, *supra* note 151, at 323-24.