Privacy, Confidentiality, and Data Security in the Age of Electronic Records and the Internet: Implications for “Human Subjects Review” in the Social Sciences

Wendy S. Davis  
Ph.D. candidate, Political Science  
Department of Political Science  
University of Utah

and

Samantha Eldridge  
Ph.D. candidate, Political Science  
Department of Political Science  
University of Utah

Western Political Science Association (WPSA) Conference, Portland, OR, 22-24 March 2012

Panel 18. 03 - Are Social Scientists to be Trusted? Proposed Revisions to the Federal Regulations Intended to Protect Research Participants

Abstract: With the advent and rapid expansion of the Internet, policy makers in many areas have been challenged to rethink what “privacy” means and how it might be protected. Laws and regulations seek to strengthen the protection of the physical and electronic records kept by a variety of institutions: federal agencies, educational institutions, medical facilities. Scientists, too, keep records that are similar in varying degrees to those kept by these other institutions. Awareness of the potential harm of breaches in the security of electronic records has lead to a highlighted concern for the protection of researchers’ data. While concerns with the protection of data and records has increased, traditional conceptions of “privacy” have been complicated and in some cases upended with spread of social media including practices such as email lists, blogs, and Facebook. A recent suggestion by the Office of the Secretary of Health and Human Services (HHS) and the Office of Science and Technology Policy (OSTP) would extend an approach used for protection of electronic records of medical patients -- known by the acronym, HIPAA -- to the protection of all research data (ANPRM, 2011, p. 44516). This work introduces the complex issues related to privacy, confidentiality and data security in the context of social science research.
Over the last 50 years - but particularly over the last 15 years with the advent and rapid expansion of the Internet - policy makers in many areas have been challenged to rethink what “privacy” means and how it might be protected. Laws and regulations seek to strengthen the protection of the physical and electronic records kept by a variety of institutions: federal agencies, educational institutions, medical facilities. Scientists, too, keep records that are similar in varying degrees to those kept by these other institutions. Awareness of the potential harm of breaches in the security of electronic records has lead to a heightened concern for the protection of researchers’ data, both as it is being generated and, later, stored. While concerns with the protection of data and records has increased, traditional conceptions of “privacy” have been complicated and in some cases upended with spread of social media including practices such as email lists, blogs, and Facebook. All scientists, but especially, social scientists have been drawn to the Internet, either as a means for conducting research (e.g., Internet surveys) or as a source of data.

In this paper, we explore two topics: (1) conceptions of privacy and confidentiality in relation to both traditional research practices and the use of the Internet; and (2) conceptions of and approaches to “data security.” Although these two topics are independent, we seek to highlight herein the complex interrelationships between the two. Moreover, the need to connect these topics is highlighted by a recent suggestion by the Office of the Secretary of Health and Human Services (HHS) and the Office of Science and Technology Policy (OSTP) to extend an approach used for protection of electronic records of medical patients -- known by the acronym, HIPAA -- to the protection of all research data (ANPRM, 2011, p. 44516). Our concern is that this suggestion takes an approach tailored to one sector--the privacy, confidentiality, and data security of individuals’ medical records--and extends it into another quite different realm, scientific research. Although there are surely lessons to be learned from an
examination of how data security is enacted in this realm (and we review some of those herein), the
variety of scientific research approaches, complicated further as scientists use the Internet both to
generate data and as a source of data, means that HIPAA conceptions of privacy and approaches to data
security may be, at best, ill suited to scientific research and, at worst, may stifle the potential use of the
Internet by scientists and, particularly, social scientists.

In this work, we explore the topics of privacy, confidentiality and data security in the context of
social science research with the understanding that there are not clearly defined answers for all of the
puzzles revealed. To frame the discussion, we first review how the concepts of privacy and
confidentiality are treated in human subjects protection policies in the US and then explore the evolving
expectations of privacy and confidentiality in the Internet world (from email lists to blogs). In the
second section, we give a broad overview of the laws and regulations that apply to electronic records in
different sectors of society analyzing both the policy text and conceptions of privacy embedded therein.  
In the third section, we report what is known about the most common breaches in “data security” for
electronic records in higher education with the understanding that the institution, at some level, bears the
burden of securing physical assets and digital resources of researchers. In the fourth section, we analyze
Internet research asking to what extent such research falls under human protections policy (in terms of
both the definitions of “human subjects” and “research”). In the final section we return to our concerns
about extensions of HIPAA to social science research and summarize why we think such an extension
would be highly problematic for social scientists.

I. Privacy and Confidentiality

The concepts of privacy and confidentiality in human subject protection policies in the United
States have broad historical roots and vary widely accordingly to context and environment. For both
biomedical and social science researchers, the central policy tenets of the federal human research
protection system are laid out in three policy documents: the 1979 *Belmont Report*, the 1993 *Office for Human Research Protections (OHRP) Institutional Review Board (IRB) Guidebook*, and the *Report and Recommendations of the National Bioethics Advisory Commission* (NBAC) (“Ethical and Policy Issues in Research Involving Human Participants Volume I Report and Recommendations of the National Bioethics Advisory Commission,” 2001). These documents regulate current research conducted across a wide range of federal departments and agencies (Schwartz-Shea & Yanow, 2006). However, in July 2011, the Department of HHS issued an Advance Notice of Proposed Rulemaking (ANPRM) recognizing the Internet has created new areas of research activity and indicating that the issues of privacy and confidentiality must be considered, once again, through a new lens.

In 1979, the *Belmont Report* was the first formal document to issue a guide to address ethical concerns arising from human subjects research. The report identified three fundamental principles that underlie all research involving human subjects: respect for persons, beneficence, and justice; however, the words “privacy” and “confidentiality” do not appear in the report (Schwartz-Shea & Yanow, 2006). Not until 1993 did the OHRP recognize the possibility that research may invade the privacy of individuals or result in a breach of confidentiality insomuch that “under certain circumstances, an invasion of privacy or breach of confidentiality may even present a risk of serious harm to subjects” (IRB Guidebook Chapter III: Basic IRB Review, 1993). Accordingly, the *OHRP IRB Guidebook* identifies the following:

*Privacy* can be defined in terms of having control over the extent, timing, and circumstances of sharing oneself (physically, behaviorally, or intellectually) with others.

*Confidentiality* pertains to the treatment of information that an individual has disclosed in a relationship of trust and with the expectation that it will not be divulged to others in ways that are inconsistent with the understanding of the original disclosure without permission (IRB Guidebook Chapter III: Basic IRB Review, 1993).

Recognizing that a researcher may have access to private information and records, the *IRB Guidebook* makes a further point that concerns about privacy and confidentiality should also pertain to the methods
used to obtain information about subjects or “personally identifiable records” (Chapter III: Basic IRB Review, 1993).

In such cases, the IRB handbook states that information collected as part of a research process should not be “improperly divulged” (Chapter III: Basic IRB Review, 1993). Moreover, “[w]hen appropriate, there [must be] adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data” (“Basic HHS Policy for Protection of Human Research Subjects”). The IRB Guidebook suggests the following: “substituting codes for identifiers, removing face sheets (containing such items as names and addresses) from survey instruments containing data, properly disposing of computer sheets and other papers, limiting access to identified data, impressing on the research staff the importance of confidentiality, and storing research records in locked cabinets” (Chapter III: Basic IRB Review, 1993). These are recommended, routine practices and are not inclusive of more elaborate procedures that may be needed to protect the identification of participants.

In the 2001 NBAC report, the definitions of privacy and confidentiality are further clarified ("Ethical and Policy Issues in Research Involving Human Participants Volume I Report and Recommendations of the National Bioethics Advisory Commission,” 2001). The report notes that privacy flows from the principle of “respect for persons” and is defined as “persons’ interests in controlling access of others to themselves and to information about them” and that confidentiality is an extension of privacy such that “confidentiality is the protection of identifiable data through agreements between participants and investigators about who may have access to the data and how the data will be managed in order to control access” (Schwartz-Shea & Yanow, 2006). Furthermore, the report instructs IRBs to review the actual procedures of a study: “those used in handling and transmitting data, eliminating linkages of data and identifiers, storing raw data (e.g., questionnaires, records, abstract forms) and data sets, planning for long-term storage and use, including sharing the data with other investigators” ("Ethical and Policy Issues in Research Involving Human Participants Volume I Report

4
Table 1 represents the differences in operationalized definitions for “privacy” and “confidentiality.”

**Table 1: Definitions of Privacy and Confidentiality**

<table>
<thead>
<tr>
<th>Source, Date</th>
<th>Privacy</th>
<th>Confidentiality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belmont Report, 1979</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>IRB Guidebook, 1993</td>
<td>“Having control over the extent, timing, and circumstances of sharing oneself (physically, behaviorally, or intellectually) with others.”</td>
<td>“Pertains to the treatment of information that an individual has disclosed in a relationship of trust and with the expectation that it will not be divulged to others in ways that are inconsistent with the understanding of the original disclosure without permission.”</td>
</tr>
<tr>
<td>National Bioethics Advisory Board Committee, 2001</td>
<td>Privacy flows from the principle of “respect for persons” and is defined as “persons’ interests in controlling access of others to themselves and to information about them.”</td>
<td>Confidentiality is an extension of privacy such that “confidentiality is the protection of identifiable data through agreements between participants and investigators about who may have access to the data and how the data will be managed in order to control access.”</td>
</tr>
</tbody>
</table>

As evident from the information presented in Table 1, there have been subtle, yet meaningful, changes to the definitions of “privacy and confidentiality.” The Belmont Report is completely void of any reference to the two terms. The IRB Guidebook (last updated in 1993, nearly two decades ago) defines the two terms, but the following disclaimer from the Office of Human Research Protections (OHRP) brings into question the reliability of the definitions as guideposts to social science researchers.

The IRB Guidebook was last updated in 1993. Developments over the intervening years have made portions of the Guidebook information obsolete, while portions of the information remain valid. There is no errata document to indicate which information has been superseded. OHRP cautions users to verify the current validity of any Guidebook information before relying on the information in a program of human subjects protection (Office for Human Research Protections).
Contributions from NBAC advance the definitions, but the researcher is required to interpret the definitions and implement the principles of privacy and confidentiality based on that interpretation. Often times institutional IRBs will clarify definitions with site-specific terminology. The variance and unreliability of existing definitions is one aspect of this complex puzzle.

The Internet, as a research tool, adds another layer of complexity. With the advent and rapid expansion of the Internet, policy makers in many areas have been challenged to rethink what “privacy” and “confidentiality” mean and how it might be protected. Scientists in the biomedical field heavily influenced the requirements set by the current policies for the protection of human subjects. Technological advances in the Internet world are opening up new areas of research activity and increasing the amount and availability of information available to researchers, especially in the social sciences field.

Yet, however encouraging such advances are, there are new issues arising as researchers delve into the technological sophistication of the 21st century. In July 2011, the Office of the Secretary of the Department of HHS in coordination with the Office of Science and Technology Policy, in their issue of an ANPRM “to request comment on how to better protect human subjects who are involved in research” (ANPRM, 2011, p. 44512), acknowledged that:

The 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 (ANPRM, 2011, p. 44513).

Furthermore, ANPRM questions whether the current policies are adequate and appropriate for the protection of human subjects and calls for changes to “modernize” the current regulatory framework.

Based on language in the ANPRM, it is apparent that policy regulators are concerned over this new area of research activity. Traditional conceptions of “privacy” and “confidentiality” have also changed and are complicated, and in some cases, upended with the use of the Internet. A fundamental
issue of Internet research is determining whether a specific online venue is considered public or private (Pace and Livingston, 2005).

In Table 2, we consider the public expectation of privacy with emphasis on online environments.

**Table 2: Expectations of Privacy**

<table>
<thead>
<tr>
<th>Environment</th>
<th>Expectations of Privacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private (Home)</td>
<td>Yes</td>
</tr>
<tr>
<td>Internet</td>
<td></td>
</tr>
<tr>
<td>• HTTPS</td>
<td>Yes</td>
</tr>
<tr>
<td>• Personal Email</td>
<td>Yes</td>
</tr>
<tr>
<td>• Employer Email</td>
<td>No</td>
</tr>
<tr>
<td>• .edu/.gov/.org (authenticated)</td>
<td>Yes</td>
</tr>
<tr>
<td>• .edu/.gov/.org (unauthenticated)</td>
<td>No</td>
</tr>
<tr>
<td>• Public Comments/OpEds</td>
<td>No</td>
</tr>
<tr>
<td>• Listserv/Email newsletters</td>
<td>Uncertain</td>
</tr>
<tr>
<td>• Wiki Spaces/Personal Webpages</td>
<td>Depends</td>
</tr>
<tr>
<td>• Podcasts (YouTube)</td>
<td>No</td>
</tr>
<tr>
<td>• Chatrooms</td>
<td>Depends</td>
</tr>
<tr>
<td>• Blogs</td>
<td>No</td>
</tr>
<tr>
<td>• Social Media (Facebook, Twitter, Linkedin)</td>
<td>Uncertain</td>
</tr>
<tr>
<td>Public Space (e.g., Park, Streets, …)</td>
<td>No</td>
</tr>
</tbody>
</table>

As is evidenced by this table, “the Internet” is not a single entity, but is a conduit—a mechanism to disperse and consume electronic content. The expectation of privacy for any particular delivery mechanism is contingent upon the technology utilized, and even then, it is unclear if citizens really have an understanding of privacy settings on various websites and web services. If there are questions about online privacy and confidentiality of users, this complicates the issue for social scientists who desire to conduct Internet research and are obligated to protect both the confidentiality and privacy of research subjects.

In this section, we explored the definitions of privacy and confidentiality, and we revealed that those definitions are not clear and instructive. The expectation of privacy and confidentiality for human subjects research is further complicated when we add the Internet as a tool that scholars use to support research activities. Even as social scientists grapple with the application of unclear definitions for
“privacy” and “confidentiality,” we report in Section II that government agencies have continually, and frequently, created policy that deals with privacy, confidentiality and data protection in arenas that extend beyond human subjects research.

II. Laws, Regulations, and the Governance of Privacy and Data Protection

In this section, we explore a chronology and a continuum of data protection policies in the United States. Table 3 represents a selection of laws that have defined how various entities must protect the data of the individuals they serve.

Table 3: A Selection of Federal Laws Governing Data Protection, Security and Privacy

<table>
<thead>
<tr>
<th>Date</th>
<th>Policy/Law</th>
<th>Synopsis</th>
<th>Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>1974</td>
<td>Privacy Act of 1974</td>
<td>Requires federal agencies to protect personally identifiable data in their control</td>
<td>Federal Government</td>
</tr>
<tr>
<td>1974</td>
<td>FERPA-Family Educational Rights and Privacy Act</td>
<td>Data collected by officials must be protected, and only officials will have access to personally identifiable information of a student (and his/her parents). Once personally identifiable information is no longer needed (for audit purposes, evaluation or enforcement of Federal legal requirements), it shall be destroyed.</td>
<td>Education: Any public or private agency or institution which is the recipient of funds under any applicable (federal) program</td>
</tr>
<tr>
<td>1996</td>
<td>HIPPA-Health Insurance Portability and Accountability</td>
<td>Provides federal protections for personal health information held by covered entities and gives patients an array of rights with respect to that information.</td>
<td>Health Care: health plans, health care clearinghouses, and to any health care provider who transmits health information in electronic form</td>
</tr>
<tr>
<td>2002</td>
<td>E-Government Act of 2002</td>
<td>Provide access to electronic government information while providing protection of personal privacy of those who access online government services.</td>
<td>Federal Government</td>
</tr>
</tbody>
</table>
In Table 1, we reviewed definitions of privacy and confidentiality as these terms relate to human subjects research. For the sectors represented in Table 3, we notice an emphasis on “personally identifiable information” and a unique melding of the concepts of privacy and confidentiality. In government, education and healthcare, the focus remains on the data itself, and it is implied that in protecting the data, the organization is protecting the person who is associated to that data. In the remainder of this section, we will investigate the four policies listed in Table 3 with attention to the issues of privacy and data protection. As we explore each policy in more detail, it is evident that this policy arena is one that is highly active and continually changing. At the conclusion of this section, we consider how incremental changes in privacy policy have affected human subjects research.

**Privacy Act of 1974**

The Privacy Act of 1974, Public Law 93-579, was created to protect the privacy rights of individuals as related to the data collected by the federal government. It was one of the first federal policies to address the management and use of citizen data. The intent of the law is to safeguard an individual’s privacy, and privacy is protected through the stipulation of “four main procedural and substantive rights in personal data” (The Privacy Act of 1974, n.d.):

1) An individual may request to see any government records kept for him or her.
2) Government agencies are required to follow “fair information practices” when collecting and managing personal data.
3) Agencies are restricted in how they can share an individual’s data with other individuals or government agencies.
4) Individuals may sue the government for violating the provisions outlined in the Privacy Act of 1974.

As technology advances (the introduction of the Internet, for example), the Privacy Act continues to be amended to address new concerns regarding the protection of citizens’ personal data (Privacy Act of 1974, n.d.). A recent proposed change to the Privacy Act occurred on October 18, 2011, when Hawaii Senator Daniel Akaka (D) introduced a new amendment to the Privacy Act (S. 1732: Privacy Act
Modernization for the Information Age Act of 2011, n.d.) that intends “[t]o modernize and improve Federal privacy laws” insomuch as it seeks to update the Privacy Act (The Privacy Act of 1974, n.d.)

The introduction of this bill and the detail contained in the language of the document, highlights the continued concerns with issues related to privacy and data protection. Many elements of the bill are administrative, such as updates to definitions, modernization of procedures related to managing data, and strengthening of penalties for misuse of data. There are also substantive aspects of the bill which create organizational change. For example, the legislation calls for the creation of a new position within the Office of Budget Management (OMB): Federal Chief Privacy Officer. This non-career appointee is to “serve as the principal advisor for federal privacy matters to the Executive Office of the President, the Director of the National Security Council, the Homeland Security Council, and the Office of Science and Technology Policy” (United States. 112th Congress, S. B. 1732). Among many responsibilities outlined, the Director is to establish policies and procedures for managing security breaches. Additionally, each agency is to appoint a Chief Privacy Officer. This group of individuals, under the leadership of the Federal Privacy Officer will comprise a “Chief Privacy Officers Council” (United States. 112th Congress, S. B. 1732). The primary function of the Council is detailed in the language of the bill.

The Council shall—share, and promote the development of best practices to assure that the use of technologies sustains, and does not erode, privacy protections relating to the use, collection, and disclosure of personal information; assure that personal information contained in systems of records are handled in full compliance with fair information practices; and evaluate legislative and regulatory proposals involving collection, use and disclosure of personal information by the Federal Government (United States. 112th Congress, S. B. 1732).

The bill also seeks to expand “the investigative authority currently granted to the Department of Homeland Security Chief Privacy Officer to other agency privacy officers” (The Privacy Act of 1974, n.d.).
The Family Educational Rights and Privacy Act (FERPA)

“The Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g; 34 CFR Part 99) is a Federal law that protects the privacy of student education records” (FERPA, 2011). Originally established in 1974, this law affects any institution receiving federal funds from the US Department of Education. The law grants rights to parents “with respect to their children's education records” (FERPA, 2011). When the student reaches the age of 18, the rights transfer from the parent to the child (referred to as the “eligible student” in the policy).

The law states that parents or eligible students (those over the age of 18) maintain the right to inspect education records collected and managed by the school. If the parent or student finds “records which they believe to be inaccurate or misleading,” the individual may request that records be corrected (FERPA, 2011). In the event that there is a disagreement between the requestor and the institution regarding the content of the records, the law establishes an appeals process for the requestor.

There are conditions under which an institution can disclose student records without consent, and those conditions are detailed in the law. For example, consent is not needed for records that have been subpoenaed. The accreditation process is also exempt from the consent requirement. Additionally, student directory information may be disclosed without consent, but the law stipulates that parents and eligible students must be given “a reasonable amount of time to request that the school not disclose directory information about them” (FERPA, 2011). Institutions are required, annually, to notify parents and students of “their rights under FERPA” (FERPA, 2011).

On April 8, 2011 the Secretary of Education, Arne Duncan, proposed amendments to FERPA under a Notice of Proposed Rule Making (NPRM) filed with the Federal Register (FERPA, 2011). The US Department of Education (USDOE) Website provides two primary reasons for the proposed changes. First, the USDOE states that there is a need for increased “accountability for institutions that handle FERPA protected records” (FERPA, 2011). Additionally, it is reported that the “proposed
regulations would give states the flexibility to share data to ensure that taxpayer funds are invested wisely in effective programs” (FERPA, 2011).

With regard to data collected and managed at the institution, the following summary of changes is posted on the USDOE Website:

- Enforcement provisions of FERPA would be strengthened to ensure that every entity working with personally identifiable information from student education records is using it for authorized purposes only.
- Schools will be able to implement directory information policies that limit access to student records, preventing marketers or criminals from accessing the data.
- States can enter into research agreements on behalf of their districts to measure the success of programs, such as early childhood programs that effectively prepare kids for kindergarten.
- High school administrators can share information on student achievement to track how their graduates perform academically in college (FERPA, 2011).

On the official Federal Register Website, the rationale for changes to FERPA relates to the use of data to support state reporting needs and federal educational policy initiatives.

These proposed amendments are necessary to ensure that the Department's implementation of FERPA continues to **protect the privacy of education records**, as intended by Congress, **while allowing for the effective use of data in statewide longitudinal data systems (SLDS)** as envisioned in the America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science Act (COMPETES Act) and furthermore supported under the American Recovery and Reinvestment Act of 2009 (ARRA).

Improved access to data contained within an SLDS will **facilitate States' ability to evaluate education programs**, to build upon what works and discard what does not, to increase accountability and transparency, and to contribute to a culture of innovation and continuous improvement in education. These proposed amendments would enable **authorized representatives of State and local educational authorities, and organizations conducting studies, to use SLDS data to achieve these important outcomes while protecting privacy under FERPA** through an expansion of the requirements for written agreements and the Department's enforcement mechanisms (Family Educational Rights and Privacy, n.d.). (Emphasis Added).
There are two points worth noting regarding the FERPA NPRM. First, it appears that the federal government would like to create more access to student data. The proposal considers allowing data sharing between individual schools, their districts, the states and the federal government. This approach to federalism and education data is, perhaps, unprecedented, and it is probable that many states are unaware that proposed changes would enable data sharing with the federal government to advance federal education policy. The second, and more salient point for our purposes, is the fact that the statutes affecting protection of student data are continually being re-evaluated and are changing to reflect changes in society and for the advancement of education policy. The final changes to FERPA were posted on December 2, 2011 (Electronic Code of Federal Regulations, 2011).

Health Insurance Portability and Accountability (HIPAA)


1 Before FERPA changes (finalized in December 2011), regulations stated that a state education authority (SEA), could not re-disclose PII obtained from local education agency (LEA) to a research organization unless the SEA had separate legal authority to act on an LEA's (or other educational institution's) behalf. The finalized regulation allows for PII between local education agencies (LEAs) and state education agencies (SEAs) to be shared without the aforementioned stipulation. In fact, the regulation states that “In the event that an educational agency or institution objects to the redisclosure of PII it has provided, the State or local educational authority or agency headed by an official listed in § 99.31(a)(3) may rely instead on any independent authority it has to further disclose the information on behalf of the agency or institution.”

The new FERPA regulation is relevant to researchers. The amendment to § 99.31(a)(6) permits any of the authorities listed in § 99.31(a)(3), including state education agencies (SEA), to enter into written agreements that provide for the disclosure of PII to research organizations for studies that would benefit the educational agencies or institutions that provided the PII to the SEA or other educational authorities, whether or not the educational authority has explicit authority to act on behalf of those agencies or institutions. The Department of Education believes that this change will have benefits for education because it would reduce the administrative costs of, and reduce the barriers to, using student data, including data in Statewide Longitudinal Data System (SLDS), in order to conduct studies to improve education programs. FERPA now permits educational agencies and institutions non-consensually to disclose PII to organizations conducting studies for, or on behalf of, educational agencies and institutions to improve instruction, to administer student aid programs, or to develop, validate, or administer predictive tests (Family Educational Rights and Privacy (n.d.).

The HIPAA statute outlines detailed requirements for data protection and incorporates provisions for both physical safeguards and technical safeguards. The official Website for the US Department of Health and Human Services (HHS) summarizes four key technical safeguards that must be followed for HIPAA compliance.

1) **Access Control.** Only authorized individuals may access protected health information (PHI).
2) **Audit Controls.** A technological infrastructure must be established to control and monitor who accesses protected health information in information systems.
3) **Integrity Controls.** Protected health information may not be “improperly altered or destroyed, and this must be monitored using “electronic measures.”
4) **Transmission Security.** “A covered entity must implement technical security measures that guard against unauthorized access to e-PHI that is being transmitted over an electronic network. (“Summary of the HIPAA Security Rule,” 2011).

The four technical safeguards are a sample representation of regulations found in the HIPAA statute. The “HIPAA Administrative Simplification” document explains 45 CFR Parts 160, 162, and 164 in detail, and this document is 101 pages in length.

HIPAA currently applies primarily to the medical industry, and while it is not required, many IRBs require that HIPAA standards apply to medically-based human subjects research. One of the recommendations in the ANPRM is that all research (including social science research) should be subject to HIPAA standards.

Accordingly, we are considering mandatory standards for data security and information protection whenever data are collected, generated, stored, or used. The level of protection required by these standards would be calibrated to the level of identifiability of the information, which would be based on the standards of identifiability under the HIPAA Privacy Rule (ANPRM, 2011, p. 44516).

The relevance of HIPAA and its application to social science research will be discussed in further detail in the last section of this paper.
E-Government Act of 2002

In the early 2000s, government was seeking ways to leverage Internet technologies to engage constituents and provide “electronic government services” (“Privacy and Civil Liberties,” 2011). The E-Government Act was created to “promote the use of the Internet and electronic government services,” as well as “to provide enhanced access to Government information and services in a manner consistent with laws regarding protection of personal privacy, national security, records retention, access for persons with disabilities, and other relevant laws” (“Privacy and Civil Liberties,” 2011).

The E-Government Act also established a Chief Information Officer (CIO) within the Office of Management and Budget (OMB) (“E-Government Act of 2002,” 2011). Under the direction of the CIO and the Chief Information Officers Council, each federal agency was required to conduct a “privacy impact assessment (PIA) for all new or substantially changed technology that collects, maintains, or disseminates personally identifiable information (PII), or for a new collection of information that is collected, maintained, or disseminated using information technology (“E-Government Act of 2002,” 2011). While most would associate this policy with web-based government services, there is important content in the bill that addresses the protection of citizen data that is collected, maintained, and accessed by the various agencies of the federal government. This legislation illustrates the pattern of government responsiveness to changes in technology and data protection issues that arise as a result of those technology advances.

Pending Legislation Related to Data Security and Privacy

On July 18, 2011, Congresswoman Mary Bono-Mack introduced the “SAFE Data Act” to the US House of Representatives (United States. 112th Congress. H. R. 2577). This bill seeks “to protect consumers by requiring reasonable security policies and procedures to protect data containing personal information, and to provide for nationwide notice in the event of a security breach” (United States. 112th Congress. H. R. 2577). The bill is aimed at ensuring identity protection in e-commerce (such as
transactions related to credit cards). There are other bills pending in the 112th Congress that focus on data security and consumer protections; the SAFE Data Act serves only as a representative for the types of pending legislation that focus on this topic.

The SAFE Data Act currently states that entities governed by HIPAA are exempt from the identity protections of this proposed legislation (United States. 112th Congress. H. R. 2577). This underscores the importance of HIPAA as a regulatory act. This may also be an indicator that once a subject area (like academic research related to human subjects) functions within the confines of HIPAA that this set of regulations will likely regulate that subject area going forward. If this assumption is true, it means that critical consideration should be given to the application of HIPAA in social science research.

*Incremental Changes in Privacy Laws and the Impact on Human Subject Research*

Since the passage of the Privacy Act in 1974, the federal government has continued to re-visit the issues of data security and privacy as it relates to the federal agencies who have stewardship for collecting, maintaining, sharing and transmitting constituent data. Since the issuance of the *Belmont Report* in 1979, social scientists have seen a variety of regulations that affect human subjects research and the protection of data related to research subjects. As technology advances, the issues related to data security and privacy become more complex.

The Department of HHS determines federal regulations for data protection standards related to human subjects research. As IRBs consider a research project, they may refer to the following guideline in the current Code of Federal Regulations (CFR): “When appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data” (“Basic HHS Policy for Protection of Human Research Subjects”). While it is not codified as such, human subject research that
is in the medical field typically applies HIPAA security standards. The Common Rule\(^2\) also mentions, in broad terms, data protection and security. Furthermore, the *IRB Guidebook*, as previously defined, provides a different, yet similar, definition for the terms “privacy” and “confidentiality” (IRB Guidebook Chapter III: Basic IRB Review, 1993). As a result, the guidelines in the Common Rule, the IRB handbook and HIPAA each approach data security differently.

The ANPRM addresses this issue by stating, “It is not clear that [IRB] members have appropriate expertise regarding data protections.” The ANPRM concedes that currently, IRBs have the responsibility for assessing informational risks but that this is an “inefficient use of an IRB’s time” (ANPRM, 2011, p. 44516). The recommendation is that “[s]tandardized data protections, rather than IRB review, may be a more effective way to minimize informational risks” (ANPRM, 2011, p. 44516). As previously noted, the ANPRM advances that the privacy and security rules defined in HIPAA should define data protection standards for *all* federally funded human subjects research projects.

As the federal government and elected leaders have shown an interest in further defining protections to constituent data, it seems likely that improvements to data related to federally funded human subject research will also be addressed. The changes in technology such as the proliferation of the Internet as a research tool and the amount of data stored on electronic systems indicates that the responsibility for ensuring data protection for human subject research is beyond the scope of any IRB or any individual investigator. Changes to data protection rules related to human subject research are inevitable. At issue is whether HIPAA is a regulation that should apply to all human subject research—including social science research. In the next section, we explore the various data breaches that occurred at colleges and universities in 2011. Currently, colleges and universities are not “covered entities” subject to HIPAA regulations. While there were multiple data breaches that affected over 200,000 individuals in 2011, a majority of the data breaches did not affect records related to human subject research.

\(^2\) The Common Rule is a standard of ethics that applies to human subjects research. In 1991, the Common Rule became part of the Code of Federal Regulations and is one of the standards used by IRBs to guide practices related to federally funded research projects.
research. Requiring HIPAA compliance for all human subject research would place a tremendous burden on colleges and universities, their technical personnel, and the way that technology assets are managed.

III. Data Breaches at Colleges and Universities and the Implications for Human Subject Research

The data breaches at colleges and universities in 2011 affected tens of thousands of students, faculty, and staff members. A five-year longitudinal study designed “to provide information to companies and consumers about the nature of data breaches” found that between 2005 and 2009, educational institutions had the highest number of breaches when compared to business, the military, financial institutions, medical institutions, and state and local governments (Garrison & Ncube, 2011). Unfortunately, we have become familiar with news stories that report data breaches on campuses across the country.

From 2005-2009, “Education had the greatest number of incidents—32.63 percent” of data security breaches—more than any other industry (Garrison & Ncube, 2011). If universities are susceptible to data security breaches of physical and digital assets, what are the implications for researchers in academia whose research data resides on university servers, on personal assets (like a laptop or mobile device) and in university physical assets (like file cabinets)? Is a data breach not a condition where confidentiality has been violated? What might the implication of such a breach entail for a social science researcher who has committed to protect the confidentiality of human subjects?

In this section, we explore this topic by examining the types of data breaches that occur on college and university campuses. The Privacy Rights Clearinghouse (PRC), maintains a database of personal information of individuals that has been “compromised through data breaches” (Chronology of Data Breaches, Privacy Rights Clearinghouse, n.d.). Their searchable database allows a user to easily query incidents by year, type, and industry. This tool allows one to focus on breaches in higher
education (as opposed to another industry like banking). The focus directs attention to the frequency and types of data breaches that have occurred on college and university campuses in 2011.

Table 4: Data Breaches at US Colleges and Universities in 2011

<table>
<thead>
<tr>
<th>Type of Data Breach</th>
<th>Description</th>
<th>Number of Reports in 2011</th>
<th>Number of Individuals Affected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unintended disclosure</td>
<td>Sensitive information posted publicly on a Website, mishandled or sent to the wrong party via email, fax or mail</td>
<td>11</td>
<td>84,437</td>
</tr>
<tr>
<td>Hacking or malware</td>
<td>Electronic entry by an outside party, malware, and spyware</td>
<td>12</td>
<td>163,072</td>
</tr>
<tr>
<td>Insider</td>
<td>Someone with legitimate access intentionally breaches information - such as an employee or contractor</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>Physical loss</td>
<td>Lost, discarded, or stolen non-electronic records such as paper documents</td>
<td>7</td>
<td>2647</td>
</tr>
<tr>
<td>Portable device</td>
<td>Lost, discarded, or stolen laptop, PDA, smartphone, portable memory device, CD, hard drive, data tape, etc.</td>
<td>4</td>
<td>11,634</td>
</tr>
</tbody>
</table>

Unintended disclosure

Oftentimes faculty and staff are completely unaware that a security breach has occurred because malicious intent was not in evidence. On June 8, 2011, a University of Mary Washington student searching for his own information on the university portal discovered that personal information of other students was readily accessible on the intranet site (Chronology of Data Breaches, Privacy Rights Clearinghouse, n.d.). The University believes that the information was accidentally posted by a faculty or staff member on May 23, 2011. “Student names, Social Security numbers, and dates of birth were

---

3 This data set does not include breaches related to patient data at medical schools.
accessible,” and three students accessed the information. (Chronology of Data Breaches, Privacy Rights Clearinghouse, n.d.). A total of 7,566 student records were affected.

**Hacking or malware (outside party)**

On January 31, 2011 St. Louis University posted a statement on their Website which acknowledged that the “University’s network was hacked on December 12, 2010” (Chronology of Data Breaches, Privacy Rights Clearinghouse, n.d.). Over 12,000 current and former employee records were compromised and personally identifiable data such as name and social security numbers were exposed in the breach. Additionally, “some students who received counseling through the University's Student Health Services may have had their names, dates of birth, tests, diagnosis and treatment information exposed” (Chronology of Data Breaches, Privacy Rights Clearinghouse, n.d.). An estimated 800 students treated at health services had their data compromised.

**Insider (party with legitimate access)**

Two former student employees at Eastern Michigan University “obtained student information and provided it to outsiders” (Chronology of Data Breaches, Privacy Rights Clearinghouse, n.d.). It is believed that the names, social security numbers, and birthdates of 45 students were compromised. The incident occurred on March 9, 2011. On October 25, 2011, one of former students was charged with eight felonies in relation to the security breach. “The charges include identity theft and using a computer to commit a crime” (Chronology of Data Breaches, Privacy Rights Clearinghouse, n.d.). The authorities have issued a warrant for the second student involved with the crime.

**Physical loss**

Incidents of physical loss are most often the result of human error. On February 3, 2011, an individual purchased a used piece of furniture at the University of Washington’s surplus store. The piece of furniture contained medical images from individuals who had been patients at the University of
Portable device

In February 2011, Albright College determined that two laptops were stolen from the financial aid office. Financial aid data is particularly sensitive because it includes parent financial information as well as student data. The data on the laptops include names, social security numbers, birthdates, and addresses. Data belonged to “faculty, staff, graduates, current and prospective students, spouses of any of these groups, and parents of students” (Chronology of Data Breaches, Privacy Rights Clearinghouse, n.d.). One laptop was recovered when a man attempted to sell it “for drug money” (Chronology of Data Breaches, Privacy Rights Clearinghouse, n.d.).

In each of the cases described, we have outlined data breaches at colleges and universities. Many of the breaches deal with employment, student, and financial aid data. What if a data breach at a college or university involves a compromise of human subject research data? There is not an organization that tracks data security breaches specifically related to research data; however, embedded in the PRC reports of educational breaches are instances of human subject research data that have been compromised. In the following two examples, the data compromised was directly related to human subject research in the medical field. Medically based research data is subject to data protections outlined HIPAA regulations. In both of these cases, proper procedure to ensure compliance with HIPPA standards was not followed—leaving the data at a high risk of exposure and breach of confidentiality for the human subjects involved.

Protecting Human Subjects: Data Breaches Related to Academic Research

In April 2011, a laptop belonging to a Tufts University research associate was stolen. The laptop contained results of research that had been conducted at Massachusetts General Hospital. Additionally,
the laptop stored a spreadsheet that contained admissions information of 73 graduate students (Chronology of Data Breaches, Privacy Rights Clearinghouse, n.d.). Names, dates of birth, and social security numbers were included in the admissions spreadsheet data. The university was notified of the laptop theft on June 16, 2011. The university reports that “the laptop was equipped with encryption software,” but the research assistant could not verify that the machine had been properly shut down; a proper shut down is required for the encryption software to work correctly (Chronology of Data Breaches, Privacy Rights Clearinghouse, n.d.).

On August 16, 2011, a physician at Indiana University School of Medicine experienced a laptop theft. Compromised information on the laptop included names, age, gender, medical record numbers, and social security numbers of multiple victims. This data was being used for research purposes (Chronology of Data Breaches, Privacy Rights Clearinghouse, n.d.). Reports state that the laptop was password protected but that the data was not encrypted (Chronology of Data Breaches, Privacy Rights Clearinghouse, n.d.).

It is not surprising that the ANPRM is considering a solution “to mandate data security and information protection standards that would apply to all research that collected, stored, analyzed or otherwise reused identifiable or potentially identifiable information” (ANPRM, 2011, p. 44525). It is clear from data breaches that occur at colleges and universities that data is in jeopardy because of technology breaches that are out of the control of the researcher. A college or university manages the technological resources of the institution and creates security policies related to electronic data sources, storage and transmission. Each individual institution regulates policy related to the use of portable devices (laptops, flash drive, etc.) and the security required for those devices. There are industry best practices for managing the security of technological resources. In some states, statutes provide guidelines for managing university technology assets. In other cases, a governing body such as a Board of Regents might
recommend security protocols. There is vast inconsistency regarding security management amongst colleges and universities in the United States. It is the recommendation of the ANPRM that HIPAA data security standards be applied to all human subject research to ensure data protection, privacy and confidentiality. In the final section of this paper, we will discuss this issue in more detail.

IV. Internet Research

The term “Internet research” is fundamentally ambiguous. It can mean retrieving data from the Internet, e.g., inspecting the websites of interest groups, universities or other groups and individuals that post information for public use. It can mean using the Internet as a means to generate evidence, that is, by surveying or interviewing individuals. And, finally, it can mean “observing” (with whatever degree of participation) human interactions in the many new venues that have emerged on the Internet, e.g., email lists, Facebook, blogs, etc. In this section of the paper, we review these three forms of Internet research and consider if each is “human subject research” and subject to IRB policy. As a final note, all Internet users make themselves vulnerable to criminal activity. Although this is beyond the scope of this section, it is relevant to IRB concerns around data security.

One quandary that social scientists face is deciphering the very definition of “human subjects research.” This issue persists regardless of if the topic is in-person research, Internet research or some other method of engagement. When an investigator conducts human subject research, the result is that information or data is obtained through intervention or interaction with the individual (45 CFR §46.102). Additionally, this data may contain identifiable private information (45 CFR §46.102). Furthermore,

**Intervention** includes both physical procedures by which data are gathered (for example, venipuncture) and manipulations of the subject or the subject's environment that are performed for research purposes [e.g., psychologists laboratories].
**Interaction** includes communication or interpersonal contact between investigator and subject. Private information includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record). Private information must be individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects. (45 CFR §46.102.)

If we utilize this definition from the Code of Federal Regulations, it would seem that the first definition of Internet research, inspection of public websites, would not require a researcher to obtain IRB approval. Review and analysis of public posted content on the Internet is not something that is unique to researchers, and it can be assumed that those persons and entities who post in the public domain do so willingly with the presumption that content will be consumed.

Researchers have long used surveys as a tool to query individuals and generate data for research questions, and the Internet simply provides a new mechanism for delivering surveys and collecting electronic responses. Research methods that incorporate surveying individuals typically require IRB approval, and the fact that a survey is administered electronically does not change the need for IRB oversight. When conducting survey research, it is imperative that the researcher consider the guidelines related to “respect for persons.” The IRB Guidebook requires informed consent and contains three main elements: information, comprehension, and voluntariness (1993). HHS has made the requirements for informed consent very clear, insomuch that “no investigator may involve a human being as a subject in research covered by this policy unless the investigator has obtained the legally effective informed consent of the subject or the subject's legally authorized representative” (45 CFR 46.116)

It is yet unclear under what circumstances informed consent is needed for Internet research. For instance, an IRB may waive the requirement for a signed consent form if it finds:

That the only record linking the subject and the research would be the consent document and the principal risk would be potential harm resulting from a breach of confidentiality. Each subject will be asked whether the subject wants documentation linking the subject with the research, and the subject's wishes will govern; or that the research presents no more than minimal risk of harm
to subjects and involves no procedures for which written consent is normally required outside of the research context. (45 CFR 46.117).

With respect to Internet research, policy regulators will need to determine whether minimal risk applies, particularly for research involving surveys, questionnaires, or similar studies. If “the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests,” then there may not be a need to monitor issues related to risk (45 CFR 46.102).

As IRB policy continues to be affected by technological changes, it is important to err on the side of caution and require that social science researchers who choose to administer online surveys as part of a research project be required to attach a consent statement. This is a prudent measure to protect both the human subject and the researcher.

Online consent presents a unique set of challenges that must be acknowledged. In an online setting, subjects agree to participate in the study by usually clicking an “accept” button on a web page (Walther, 2002). Issues with informed consent can occur when researchers intentionally misinform or do not fully disclose relevant information to participants (Frankel & Siang, 1999). The use of pseudonyms further complicates the issue. In this case, since the identity of the participant cannot actually be determined, it cannot be known if legal consent was given (Pace & Livingston, 2005). Policy regulators will need to work toward a standardized measure to validate the consent of a participant who is the subject of Internet research.

When a researcher observes behavior online, is this human subject research? We make the contention that it is, and as such, this type of research method is subject to IRB oversight. It may be assumed that public behavior is “fair game” for observational research; however, human research subjects may not be aware that their public behavior is being portrayed to a wider audience of complete strangers (Shrag, 2008). The IRB handbook describes this as “covert observation.” The Internet presents an opportunity for observance where one previously did not exist, and the current CFR does not
account for online observational methods. According to Shrag, this may call into question the adequacy of current research guidelines and their ability to account for changes in technology and its application in social science research (p. 144).

Researchers using the Internet for research purposes may use pseudonyms or assume anonymous identities online. Researchers who use social network sites such as Twitter and Facebook sometimes create fake accounts to extract profile information. Information is obtained by “lurking” or subscribing to mailing lists and visiting communities without actually participating in the list or community and by “harvesting” or extracting information from online communication for research purposes (Berry, 2004). Walther (2002) suggests that the Internet provides varying degrees of anonymity and promotes the misrepresentation of identity (by those being observed) with respect to gender, age, and geographic location. The ability to assume or fabricate identity complicates matters regarding informed consent, correct identification of research participants, and protection of underage subjects (Berry, 2004). For those studies that require it, “ascertaining the true identity of research participants requires the collection of personally identifiable information, yet such information poses risk to individual privacy” (Pace & Livingston, 2005).

The emergence of online research presents several conundrums for the researcher and the IRB. In a recent study seeking to understand the state of the US Institutional Review Board system, Buchanan and Ess (2009) found that IRBs generally do not know what protections should apply to online research. Buchanan and Ess surveyed 334 respondents inquiring on their institutions’ Institutional Review Board’s experience with Internet research. Sixty-two percent of the respondents did not have guidelines or checklists in place for reviewing Internet research-based protocols. Seventy-four percent did not provide specific training related to Internet research issues, and nearly all surveyed have found Internet research an area of concern or importance.
Buchanan and Ess concluded that in light of the ever-changing nature of the Internet, research ethics should adapt accordingly. However, Buchanan and Ess (p. 45) found that IRBs often ignore the complexities of such research and thereby risk harming subjects. Alternately, they found that IRBs apply such restrictive models that they inhibit researchers from pursuing important online endeavors (p. 45). It is important that social scientists and policy regulators have an understanding of the key concerns related to risk before undertaking Internet research so that they can be proactive in addressing ethical issues before they emerge (Ess, 2001 as cited in Berry, 2004, p. 326). If ANPRM wishes to call upon researchers to “more effectively guide social policy and practices” (ANPRM, 2011, p. 44513), it must recognize the differences posed by research in the biomedical field compared to social and behavioral science research. The discussion of Internet research in this section serves to illustrate the complex nature of using emerging technologies in human subject research. Additionally, this section (using Internet research as an example) highlights that the proposed application of biomedical regulations as the standard for social scientists does not necessarily make sense.

V. Government Regulation of Privacy, Confidentiality and Security for Human Subject Research

When commissioners met at the Belmont House in 1976, they deliberated over the role of IRBs. Stephen Toulmin, a participant in the Belmont discussions, said:

One of the questions that is obviously going to have to be looked at very carefully when we move on to the IRB examination phase is simply the question of whether this same set of procedures can meaningfully and effectively be employed for the appraisal of biomedical research involving human subjects and these other kinds of research, especially in the field of social science (quoted in Schrag, 2010, p. 82).

It is somewhat ironic that thirty-five years later, the same discussion point persists. Can the same set of procedures employed for biomedical research apply to the field of social science?

As we have noted, there are serious concerns regarding data security and privacy of records that are under the purview of the federal government. Recent changes to the Privacy Act and FERPA indicate that the federal government is interested in maintaining protections and privacy for data
submitted by individuals and held within the stewardship of various government agencies. Federally funded projects that entail human subjects also fall within the realm of government protections.

Currently, the CFR provides ambiguous guidelines for IRBs and researchers regarding the topic of data security and data protections. The ANPRM cites the Common Rule and HIPAA as two regulations, which are “overlapping and sometimes inconsistent…causing confusion and frustration among investigators, IRBs, and others trying to comply with both sets of requirements.” One analysis comparing the two regulations states, “Because the HIPAA Privacy Rule and the “Common Rule,” evolved in different contexts, [they] take different approaches to protecting privacy” (Pritts, 2008). The definition on the Department of HHS web site clearly indicates that HIPAA applies to Personally Identifiable Information (PII) in the healthcare industry.

A major concern for social science research is that HIPAA data protection standards originated in the medical field and concern the protection of patient data. Within the Department of Health and Human Services, the Office of Civil Rights “administers and enforces” the HIPAA Privacy Rule and HIPAA Security Rule (HIPAA Administrative Simplification Statute and Rule, n.d.). The Centers for Medicare & Medicaid Services administer and enforce: Transaction and Code Sets Standards, Employer Identifier Standards, and National Provider Identifier Standards (HIPAA Administrative Simplification Statute and Rule, n.d.). The HIPAA Administrative “Simplification” Regulation Text is a 101-page document that “summarizes” the 45 CFR Parts 160, 162, and 164. It is difficult to accept that the application of HIPAA standards to social science research will “reduce burden, delay, and ambiguity for investigators” (ANPRM, 2011, p. 44514). It is probable that a separate regulatory board or expert would have to review social science research proposals to ensure HIPAA compliance. The regulatory code is difficult for a non-medical principal investigator to consume. It could be considered a “burden” for a researcher to ensure HIPAA compliance is met. It would be unfortunate if research projects were delayed or rejected because of the application of HIPAA to social science research projects.
Even without conducting a full analysis of HIPAA regulations, it is obvious that HIPAA addresses the primary issues stated in the ANPRM: privacy, confidentiality and data protection. It would seem that the goal in establishing HIPAA as a standard for all researchers is an attempt at consistency and uniformity. Interestingly, the ANPRM acknowledges the following paradox:

Many critics see little evidence that most IRB review of social and behavioral research effectively does much to protect research subjects from psychological or informational risks. **Over-regulating social and behavioral research in general may serve to distract attention from attempts to identify those social and behavioral research studies that do pose threats to the welfare of subjects and thus do merit significant oversight** (ANPRM, 2011, p. 44513). (Emphasis Added.)

Writing for the *Northwestern Law Review*, a collaborative team of authors state, “Social science research paradigms fit poorly into the thrust of medically-driven IRB protocol templates and language” (Beldsoe et al., 2007).

Is this the proper direction for social science research? Could the HIPAA requirement (with its noted complexities) become an impediment for researchers? Does the invocation of HIPAA create barriers to creativity and freedom of research? Was the HIPAA recommendation the choice because it already applies to much of biomedical research?

The Consortium of Social Science Associations (COSSA) issued a formal white paper response to the ANPRM in July 2011. In the response, the authors acknowledge that it is understandable why some believe that adapting HIPAA to “to meet the needs of data security and data protection” across different spheres can provide the “right foundation” (p. 42). However, in a strongly worded statement, COSSA members wrote the following:

We have grave reservations, however, about the use of HIPAA as model for privacy protection and data security practices under the Common Rule. HIPAA has a very specific purpose—to alleviate privacy concerns about identifiable information in administrative health records. Hence, in dealing with research uses of health data, its central concern is ensuring that data sets are fully or partially deidentified, and regardless of use it is concerned with preventing data breaches and redressing any that seriously threaten a person’s interest (p. 42).

The COSSA white paper recommends that “instead of turning to HIPAA, the drafters revising
45CFR46 (1) commit to examining options available for researchers establishing data protection plans and for oversight of those plans in ways that are efficient, flexible, and in accord with the data being collected, (2) plan to consult experts for guidance” (p. 45).

In considering how HIPAA will affect social science research, it is also important to consider the impact to the research institutions that must enforce HIPAA as a new standard for all human subject research. Currently, colleges and universities are not subject to federal HIPAA regulations. There has been great debate in the past ten years regarding personnel data and the applicability of HIPAA. The regulation clearly states that employment data is not subject to the current rules. Applying HIPAA to human subject research projects at universities would completely alter accountability for an institution with respect to data security. It would become incumbent upon colleges and universities who manage federally funded research projects to incorporate detailed HIPAA requirements into existing institutional security policies.

What might the implication be for colleges and universities if HIPAA is employed as the regulation that governs all federally funded research? Most of the technology related aspects of HIPAA (such as the mandate to utilize encryption software) are out of the control of the individual researcher. How much of this burden is assumed by the institution, and what does that mean for researchers in the short-term and in the long term? Does data remain property of the institution (which then must safeguard and protect it), or does research data remain under the stewardship of the researcher? What does this mean for institutions who do not have the financial resources to become HIPAA compliant? The answers to these questions are quite important in determining the implications for data security in social science research.

One must ask if the proposed changes meet the objective to “enhance the effectiveness of the research oversight system by improving protections for human subjects while also reducing burdens, delays and ambiguity for investigators and research subjects” (ANPRM, 2011, p. 44514)? If not, what
type of mandated data security protection standards will ensure that human subject research is protected? The ANPRM does not offer an alternative to HIPAA as a regulation option.

It is our opinion that while data protection, privacy and confidentiality are vitally important in social science human subject research, HIPAA is specifically designed to protect medical data. Extending the HIPAA regulation to social science research would most likely not reduce the burden for investigators and research subjects; it is likely that burdens would be increased for investigators and the colleges and universities who would have the responsibility to ensure HIPAA compliance for all federally funded human subject research projects.

Moreover, we argue that the definitions of privacy, confidentiality and data protections must be flexible enough to account for the emergence of the Internet as a research tool. New privacy and data security regulations must consider various Internet research approaches such as electronic surveys, interviews and online observation. The regulations should be flexible to account for new and emerging web-based technologies that might support the activities of social science researchers in the future.

As we stated at the beginning of this work, we are presented with a complex set of puzzles that challenges current definitions and future considerations for researchers. The intent of this paper was to reveal the complexity of the issues and posit questions to stimulate consideration of these issues. Unquestionably, this is the beginning of a conversation that will likely develop as regulations are codified and as “research” continues to be redefined by the still emerging technological age.
References


