Legal Issues in the Use of Electronic Data Systems for Social Science Research

John Petrila, J.D., LL.M.

College of Behavioral
and Community Sciences
University of South Florida

Contents

Introduction

This paper provides an overview of legal issues in using and linking large datasets for social science research. The paper is based on three assumptions. First, linked datasets are essential in conducting services research and policy analyses. Second, it is usually legally possible to collect information and create and link data, though the legal rules for different categories of information may vary. Third, while privacy and confidentiality laws are critical in thinking about these issues, the legal rules governing the security of data are as important.

The paper has four sections. The first briefly summarizes federal and state laws that affect the privacy of several types of information that social science researchers may wish to access. The second discusses Institutional Review Boards (IRBs), including important proposed changes in the federal rules governing IRBs. The third summarizes the relevant law on the security of electronic health data, primarily that found in the regulations implementing the Health Insurance Portability and Accountability Act (HIPAA). The paper concludes with a brief discussion of enforcement and penalties for violating confidentiality laws.

Space does not permit an exhaustive discussion of these issues, and throughout, the reader is referred to other resources that provide more detailed information regarding discrete topics. The paper should also be read in conjunction with other papers commissioned as part of this series, specifically the paper titled *Ethical Use of Administrative Data for Research Purposes* by Stiles and Boothroyd and the paper on technology entitled *An Overview of Architectures and Techniques for IDS Implementation.* Finally, this paper is not a substitute for legal counsel and readers should always consult their own counsel when legal advice on a specific issue is required.

Relevant Laws on Privacy and Confidentiality

INTRODUCTION

Large, electronic datasets are an increasingly important tool in social science research. These datasets may contain information from one or more sources, including healthcare records, criminal justice and juvenile justice records, education records, child welfare records, or judicial records. While researchers may choose to use a single dataset in their research, linking datasets to maximize the amount of available information is increasingly common.

While large, linked datasets are a boon to research, they raise concerns for individual privacy and confidentiality. First, the datasets may contain information that identifies individuals and things about them (health needs, financial status, involvement with the justice system) that the individual may wish to keep private. Second, linking discrete datasets may compound the amount of information revealed about the individual, far beyond a lone dataset. For example, one group of commentators asserts:

> Many questions require linking micro level survey or census data with spatially explicit data that characterize the social, economic, and biophysical context in which survey or census respondents live, work, and/or engage in leisure activities. Once the precise spatial locations of a person's activities are known, these locations serve as identifiers that can be used as links to a vast array of spatial and social data. This linkage poses challenges to issues of confidentiality, data sharing among scientists, and archiving data for future scientific generations (VanWey, Rindfuss, Gutmann, Entwisle, & Balk, 2005).

Myriad legal rules, sometimes consistent, sometimes not, govern access to individually identifiable information. Laws regulating access to different types of information vary because privacy and confidentiality statutes and regulations have emerged in different situations, usually to address particular types of information. For example, the Family Educational Rights and Privacy Act (FERPA) was enacted by Congress in 1974 to address the confidentiality of education records, while two federal statutes[1] enacted in the early 1970s were the foundation for federal regulations on the confidentiality of alcohol and substance abuse treatment records. Both education and substance use/alcohol treatment records have strict confidentiality protections, but the scope of the protections and conditions of access differ. Therefore, the legal conditions under which information will be accessible for research depends on the source, type, and location of the information.

In addition, federal and state laws permitting access to electronic datasets for research purposes often distinguish between access to information that identifies the individual and information that does not. In such cases, the researcher must decide whether identifying information is essential to research. If not, the legal rules for accessing de-identified data may be easier to use.

[1] The two statutes are the Comprehensive Alcohol Abuse and Alcoholism Prevention, Treatment and Rehabilitation Act of 1970 and the Drug Abuse Prevention, Treatment and Rehabilitation Act of 1972.

With these introductory remarks in mind, the rest of this section provides brief summaries of a number of important federal and state laws. It first identifies a law and core definitional terms, then briefly discusses research access to protected information for that particular law. Each discussion concludes with a summary of any provisions in the specific law that address de-identified data.

FEDERAL POLICY FOR THE PROTECTION OF HUMAN SUBJECTS (THE COMMON RULE)

Because it establishes the basic framework that regulates most social science research, the discussion begins with the Federal Policy for the Protection of Human Subjects, also known as the Common Rule. The Common Rule is a federal policy designed to protect human subjects, first published in 1991. It is called the Common Rule because 15 federal agencies agreed to be bound by its terms, in research involving human subjects conducted, supported, or regulated by the agencies. A description of the Common Rule prepared by the Office for Human Research Protections (OHRP) of the United States Department of Health and Human Services, including links to the federal agencies that have agreed to be bound by it, can be found here: http://www.hhs.gov/ohrp/humansubjects/commonrule/index.html.

The Common Rule defines research as "a systematic investigation, including development, testing, and evaluation, designed to develop or contribute to generalizable knowledge" (45 C.F.R. § 164.501.).[2] In general, research relying on linked datasets containing individually identifiable information will require approval from an Institutional Review Board, the entity charged with protecting human subjects. To approve a research project, the IRB must find that the researcher meets a list of criteria (discussed in more detail below in Research Approval). One in particular is worth noting here. The IRB must assure "when appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data" (46.111(a) (7)). The IRB must consider this for all studies under its review. However, when the researcher is using information protected by the Health Insurance Portability and Accountability Act (HIPAA), then an additional set of rules (discussed below) will apply.

[2] The Common Rule exempts certain types of research. This includes research conducted in educational settings, involving normal educational practices; research involving the use of educational tests; research involving the collection or study of existing data, documents, records, or pathological or diagnostic specimens where the subject cannot be identified; and certain research and demonstration projects examining public benefit or service programs (46.101(b)(1-6). The regulation itself should be read for the precise wording of these exemptions.

The federal government has proposed amendments to the Common Rule designed to strengthen human subjects protections. Among other things, the amendments would establish mandatory data security and information protection standards for research using identifiable or potentially identifiable data. In addition, IRB approval would be required for all studies conducted at institutions receiving funding from Common Rule agencies, rather than only those studies funded by a Common Rule agency. More information about the proposed amendments is provided in Research Approval, on IRBs, and the proposed amendments are discussed in considerable detail here: http://www.hhs.gov/ohrp/humansubjects/anprm2011page.html.

THE PRIVACY ACT OF 1974

This federal statute establishes rules for protecting records about individuals maintained by federal agencies containing "personally identifiable information" (or PII). PII includes but is not limited to " … education, financial transactions, medical history, and criminal or employment history and that contains … name, or the identifying number, symbol, or other identifying particular assigned to the individual, such as a finger or voice print or a photograph" (5 USC 552a(a)(4)). The Privacy Act was enacted because of concerns over intrusions into personal information maintained in computerized "systems of records." In 1988, Congress enacted the Data Matching and Privacy Protection Act of 1998 to supplement the Privacy Act. This act requires federal agencies that use computer matching activities to create Data Integrity Boards. In addition, Privacy Act provisions create rules that limit the ability of agencies to run matching programs on systems of records, absent a written agreement between the agencies. The matching agreement must state:

- the purpose and legal authority for conducting the matching program;
- the justification for the program and its anticipated results, including an estimate of any savings;
- a description of the records that will be matched, including each data element used, the approximate number of records to be matched, and the projected starting and completion dates of the matching program;
- various procedures for: giving notice to potentially affected individuals; verifying the accuracy of the program's results; keeping the records current and secure; and regulating the use of the results;

- any assessments of the accuracy of the records to be used; and
- a section allowing the Comptroller General access to all of the records it deems necessary in order to monitor compliance with the agreement.

Agencies provide information regarding their matching requirements. For example, the UNITED STATES Department of Education's requirements can be found here: http://www2.ed.gov/policy/gen/leg/foia/acsom6105.pdf

The Privacy Act and Researcher Access to PHI

The Privacy Act has stringent confidentiality provisions but permits disclosure without the subject's consent for a "routine use," defined as "the use of such record for a purpose which is compatible with the purpose for which it was collected" (5 USC 522a (a) (7)). This has been used to permit researcher access even to identifiable data. An example of how Medicare data, which is protected by the Privacy Act, can be accessed by researchers is provided under "Research Issues" in http://www.resdac.org/medicare/medicarefaq.asp. Note also that if a HIPAA "covered entity" (discussed immediately below) has the data, then HIPAA rules also apply.

HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPAA)

HIPAA creates rules for "protected health information" (PHI) in the control of a "covered entity"." PHI is defined as

> … any information, whether oral or recorded in any form or medium that is created or received by a health care provider, health plan, public health authority, employer, life insurer, school or university, or health care clearinghouse *and* relates to the past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual. (45 CFR 160.103)

A "covered entity" is either a "health plan"[3], a "health care provider" that transmits information in electronic form in connection with a HIPAA transaction[4], or a

[3] Health plans include health insurance companies, government plans paying for health care such as Medicare or Medicaid, and HMOs.

[4] Health care providers include individual health care professionals as well as entities such as hospitals or nursing homes.

"health care clearinghouse."[5] (For more information regarding covered entities, see guidance from the United States Department of Health and Human Services at http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/index.html). It is worth noting that HIPAA exempts from its coverage education records and treatment records maintained in student health records that meet the definition of "education records" within the Federal Education Rights and Privacy Act (FERPA, discussed immediately below). The Departments of HHS and Education have issued a joint guidance on the relationship between FERPA and HIPAA which can be found here: http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/hipaaferpajointguide.pdf.

HIPAA and Researcher Access to PHI

HIPAA uses the same definition of "research" as that found in the Common Rule. As discussed in more detail below, it also provides for various ways to access PHI held by a covered entity. If the information qualifies as PHI, it may be released for research purposes in the following circumstances (45 C.F.R. § 164.512(i)):

- If the subject of the PHI has granted specific written permission through an Authorization that satisfies section 164.508;
- For reviews preparatory to research with representations obtained from the researcher that satisfy section 164.512(i) (1) (ii) of the Privacy Rule. This may include activities associated with recruitment of subjects;[6]

[5] Health care clearinghouses include those that standardize health information, such as a billing service that processes or facilitates the processing of data from one format into a standardized billing format.

[6] Activities that would qualify as "work preparatory to research" include identification of potential research subjects and preparation of a research protocol. The principle of minimal necessity applies, which means that the researcher should access PHI only as necessary to do this preparatory work. There are certain conditions for access in this case. The researcher cannot remove PHI from the site and must represent to the custodian of the PHI that:

• The use or disclosure is sought solely to review PHI as necessary to prepare the research; protocol or other similar preparatory purposes;

• No PHI will be removed from the covered entity during the review; and

• The PHI the researcher seeks to use or access is necessary for the research purposes.

Researchers who use linked electronic data often do not recruit research subjects. For those interested in the topic, which involves both identifying and contacting potential research subjects, the National Institutes of Health provides guidance, as part of its more general discussion of the Privacy Rule and research (http://privacyruleandresearch.nih.gov/clin_research.asp).

- For research solely on decedents' information with certain representations and, if requested, documentation obtained from the researcher that satisfies section 164.512(i)(1)(iii) of the Privacy Rule;[7]
- If the covered entity receives appropriate documentation that an IRB or a Privacy Board has granted a waiver of the Authorization requirement that satisfies section 164.512(i) (discussed in more detail in Research Approval below);
- If the covered entity obtains documentation of an IRB or Privacy Board's alteration of the Authorization requirement as well as the altered Authorization from the individual;
- If the PHI has been de-identified in accordance with the standards set by the Privacy Rule at section 164.514(a)-(c) (in which case, the health information is no longer PHI) (discussed in more detail immediately below);
- If the information is released in the form of a limited dataset, with certain identifiers removed and with a data use agreement between the researcher and the covered entity, as specified under section 164.514(e) (discussed in more detail below).

For a discussion of clinical research and the Privacy Rule prepared by the National Institutes of Health, see http://privacyruleandresearch.nih.gov/pr_02.asp. For a discussion by the United States Department of Health and Human Services Office for Civil Rights, the primary enforcer of HIPAA standards, see http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/research.html.

Alternatives to Use of PHI Under HIPAA

HIPAA provides alternatives to accessing health information that does not identify the individual through **de-identified data**. HIPAA provides for two ways of de-identifying data, one through stripping data elements, the other through statistical analysis.

[7] A researcher may wish to use PHI from decedents. To do so, the researcher must represent to the covered entity that the PHI is necessary for research and that the research will involve decedents, and not family members or others. The covered entity also may request documentation of death. If these conditions are met, authorization or waiver by an IRB or Privacy Board is unnecessary.

To accomplish the former, the following items must be deleted from a person's PHI:

1. Names
2. Geographic subdivisions smaller than a state
3. Dates (except year) directly related to patient
4. Telephone numbers
5. Fax numbers
6. E-mail addresses
7. Social security numbers
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers
13. Device identifiers and serial numbers
14. Web URLs
15. Internet Protocol (IP) address numbers
16. Biometric identifiers, including finger and voice prints
17. Full face photographic images and any comparable images
18. Any other unique identifying number, characteristic, or code, except as permitted under HIPAA to re-identify data.

As an alternative to removal of the 18 elements for de-identification, "a person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable" may determine that there is a "very small" risk that the information could be used to identify any individuals from the data, alone or in combination with other reasonably available information. The United States Department of Health and Human Services has provided guidance on the topic of statistical de-identification (HHS 2005; also see http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/deidentificationworkshop2010.html)

Researchers may also use a **limited dataset** under the HIPAA Privacy Rule as an alternative to using PHI. A limited dataset excludes many of the same data elements

as a de-identified dataset but permits inclusion of all dates related to the patient, five-digit zip codes, and city as indirect identifiers.

A limited dataset is PHI that excludes "direct identifiers" of the individual, relatives of the individual, employers or household members.

Specifically excluded are:

1. Name
2. Postal address other than city, town, state and zip code
3. Telephone numbers
4. Fax numbers
5. E-mail address
6. Social security number
7. Medical record number
8. Health plan beneficiary identifiers
9. Account numbers
10. Certificate/license numbers
11. Device identifiers and serial numbers
12. Web URL
13. Internet protocol (IP) address numbers
14. Biometric identifiers including finger and voice prints
15. Full face photographic images, and
16. Any other number, characteristic, or code that could be used to identify the individual.

The researcher and covered entity providing the dataset must sign a Data Use Agreement that (1) describes the permitted uses and disclosures of the information and (2) prohibits any attempt to re-identify or contact the individuals. An example of a Data Use Agreement for a limited dataset can be found here http://info.dhhs.state.nc.us/olm/manuals/dhs/pol-80/man/DHHS_Data_Use_Agreement_Template.pdf.

Note finally that a question may be raised regarding whether a researcher accessing data from a covered entity must sign a "business associate" agreement. A "business associate" is a person or entity that performs functions or activities that involve the use or disclosure of PHI on behalf of a covered entity or provides services to

a covered entity. Examples include, but are not limited to, third parties that do claim processing for the covered entity; provide accounting services; act as counsel; or provide utilization reviews. A covered entity must have a "business associate" agreement with third parties performing business associate activities. However, the United States Department of Health and Human Services' Office for Civil Rights (OCR) is clear that researchers accessing PHI with authorization, pursuant to a waiver of authorization, or through a limited data set, do *not* require a business associate agreement (see http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/businessassociates.html, Other Situations in Which a Business Associate Contract is NOT Required). According to OCR, the research is not an activity such as payment or health care operations, and therefore the requirements do not apply. For an example of one university's explanation of why it does not enter business associate agreements for research see http://www.hpitp.buffalo.edu/hipaa/Research/DataExtraction.htm. Therefore, as long as researchers are not performing activities that require a business associate agreement, such an agreement is not required by HIPAA.

FEDERAL EDUCATION RIGHTS AND PRIVACY ACT (FERPA)

FERPA regulates the confidentiality of education records. FERPA, administered by the United States Department of Education (ED), defines education records broadly as those records directly related to a student and maintained by an educational agency or institution or by a party acting for the agency or institution (34 CFR 99.3). (For the text of FERPA see 20 USC 1232g and 34 CFR Part 99, found at http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&tpl=/ecfrbrowse/Title34/34cfr99_main_02.tpl; Resources and guidance on FERPA provided by ED can be found at http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html).

Written consent of the parent of a minor student is generally required prior to release of "personally identifiable information" (PII) about the student. Under FERPA, PII includes but is not limited to:

a. The student's name
b. The name of the student's parents or other family members
c. The address of the student or student's family
d. A personal identifier, such as a social security number, student number, or a biometric record: a record of one or more measurable biological or behavioral characteristics that can be used for automated recognition of an individual, i.e., fingerprints, handwriting, etc.

e. A list of personal characteristics

f. Other information that would make a student's identity easily traceable

Some types of information are not considered part of the education record, for example, treatment records maintained in connection with the treatment of students 18 or older (99.3b). Such records, if created by a covered entity and containing PHI, would be covered by HIPAA (or a more stringent state confidentiality law if available. See the discussion of preemption below.) On the other hand, notes entered in a student's record, for example, by a school nurse, would be considered part of the education record and subject to FERPA, rather than HIPAA. This potentially confusing situation is addressed by joint guidance on the relationship between FERPA and HIPAA issued by the United States Departments of HHS and Education found here http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/hipaaferpajointguide.pdf.

FERPA and Researcher Access to Education Records

Education records may be released to an "authorized representative" of the educational institution or agency for audits, evaluation, or enforcement or compliance activities related to educational activities. PII also may be released without written consent to organizations conducting studies for, or on behalf of, the educational institution for test development and validation, the administration of student aid programs, or to improve instruction. A written agreement between the educational institution and organization conducting the study is required, and the agreement must detail how the study will be conducted, restrict the use of PII only for study purposes, and assure that no parent or student will be personally identified as a result of the study (99.31(a)(6). The U.S. Department of Education has a website devoted to the protection of human subjects at http://www2.ed.gov/about/offices/list/ocfo/humansub.html, and provides a sample agreement between an educational institution and an authorized representative in Appendix B at http://www2.ed.gov/about/offices/list/ovae/pi/cte/uiferpa.html.

In many jurisdictions, education records have been difficult to access for research because of federal regulatory rules that define an "authorized representative" as an entity over which the institution exercises direct control. However, the United States Department of Education has recently promulgated new rules that may ease these difficulties. In originally proposing the new rules, the Department concluded that the "direct control" requirement was overly restrictive and prohibited educational institutions from providing information even to other state agencies for purposes

of conducting analyses directly relevant to the relationship between education and other programs. In proposing less restrictive regulations, ED stated, "We do not believe such a restrictive interpretation is warranted *given Congress' intent in the [American Recovery and Reinvestment Act] to have states link data across sectors"* (emphasis supplied). Accordingly, the regulation now permits an educational institution or agency to designate an "authorized representative" to perform various functions, including program evaluation, but no longer requires the representative to be under the "direct control" of the agency. For the regulation with commentary by DE see http://www.gpo.gov/fdsys/pkg/FR-2011-12-02/pdf/2011-30683.pdf. What is particularly important is that ED is

> "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."

ED also makes clear that

> These final regulations allow FERPA-permitted entities to disclose PII from education records without consent to authorized representatives, which may include other state agencies, or to house data in a common state data system, such as a data warehouse administered by a central state authority for the purposes of conducting audits or evaluations of federal- or state-supported education programs, or for enforcement of and ensuring compliance with Federal legal requirements relating to federal- and state-supported education programs (consistent with FERPA and other Federal and State confidentiality and privacy provisions). (See http://www.gpo.gov/fdsys/pkg/FR-2011-12-02/pdf/2011-30683.pdf at 75637).

Given the difficulty many researchers have had in accessing education records, the amendments should prove helpful.

The new rule continues to require a written agreement between the agency and its authorized representative regarding conditions under which education data would be accessed, used, and ultimately returned or destroyed (34 CFR 99.3).

Use of De-Identified Education Records Under FERPA

FERPA (99.31(b) provides for the de-identification and release of education records if certain requirements are met. Specifically, the educational institution must remove all personally identifiable information (PII). A code must be attached to each record that may allow the recipient of the information to match information. However, the educational agency or institution cannot reveal how it generated or assigned the code and the code can be used for no purpose other than identifying the de-identified record. Finally the code cannot be based on the student's social security number or other personal information.

The National Center for Education Statistics has published an excellent guide to privacy and confidentiality of education records, including de-identified records, titled SLDS Technical Brief: Guidance for Statewide Longitudinal Data Systems: Basic Concepts and Definitions for Privacy of Student Education Records. It may be found here http://nces.ed.gov/pubs2011/2011601.pdf.

FEDERAL REGULATIONS GOVERNING THE CONFIDENTIALITY OF ALCOHOL AND SUBSTANCE ABUSE TREATMENT RECORDS

A very stringent federal regulation (often referred to as 42 CFR Part 2) protects the confidentiality of drug and alcohol treatment records. This regulation governs the disclosure of "any information, whether recorded or not, relating to a patient received or acquired by a federally assisted alcohol or drug program" (42 CFR 2.11). Unlike HIPAA, which covers PHI only when it is under the control of a covered entity, 42 CFR protections follow records regardless of who has possession. For example, if a court orders disclosure of information under 42 CFR, the court and parties to the proceeding continue to be bound by the requirements of 42 CFR even though they are not "federally assisted programs."[8] However, other types of records that identify individuals as being in treatment for alcohol or substance abuse (for

[8] A program is "federally assisted" if it: (1) is conducted entirely or in part by any federal agency or department (with some exceptions for Veterans Administration and Armed Forces programs); (2) is conducted under a license, certificate, registration, or other authorization from any federal agency or department, including certified Medicare providers, authorized methadone maintenance treatment providers, and programs registered under the Controlled Substances Act to dispense controlled substances for alcohol or drug abuse treatment; (3) is tax-exempt or to whom contributions are tax deductible; or (4) is the recipient of any federal funds. 42 C.F.R. § 2.12(b).

example, those created by law enforcement) are not governed by 42 CFR because the records were not generated by a federally assisted program.

42 CFR and Researcher Access to Alcohol and Substance Abuse Treatment Records

Despite its restrictiveness, the regulation does permit the use of covered information for research without the person's consent, if the federally assisted program director finds that the following requirements are met (42 CFR 2.52). Specifically, the recipient of the information:

1. Is qualified to conduct the research;
2. Has a research protocol under which the patient identifying information will meet the security requirements of § 2.16 of these regulations (or more stringent requirements); and will not be re-disclosed except as permitted by 42 CFR Part 2 and
3. Has provided a satisfactory written statement that a group of three or more individuals who are independent of the research project has reviewed the protocol and determined that the rights and welfare of patients will be adequately protected; and that the risks in disclosing patient identifying information are outweighed by the potential benefits of the research.

A person conducting research may disclose patient identifying information only back to the program from which that information was obtained and may not identify any individual patient in any report of that research or otherwise disclose patient identities. [52 FR 21809, June 9, 1987, as amended at 52.]

For a good discussion by Kamoi and Borzi comparing the HIPAA Final Rule with 42 CFR Part 2, titled "A crosswalk between the final HIPAA privacy rule and existing federal substance abuse confidentiality requirements" visit: http://www.gwumc.edu/sphhs/departments/healthpolicy/CHPR/downloads/behavioral_health/bhib-18-19.pdf.

THE HOMELESS MANAGEMENT INFORMATION SYSTEM (HMIS)

Social science researchers also may be interested in information regarding individual experiences with homelessness. Congress directed the United States Department of Housing and Urban Development (HUD) to collect data on homelessness, which is done today through the Homeless Management Information System (HMIS, see http://portal.hud.gov/hudportal/HUD?src=/program_offices/comm_planning/homeless/hmis).

The HMIS rules protect the confidentiality of "protected personal information," or PPI. The definition of PPI is similar, though not identical, to the definitions of protected information used in the other federal laws discussed in this section. PPI is any information maintained by a homeless organization that:

- Identifies, directly or indirectly, a specific individual;
- Can be manipulated by a reasonably foreseeable method to identify an individual; or
- Can be linked with other available information to identify an individual.

See http://www.hmis.info/ for further information posted by the United States Department of Housing and Urban Development regarding the HMIS system.

Programs that collect data for an HMIS system are required to collect fifteen data elements. These are name, social security number, date of birth, race, ethnicity, gender, veteran status, disabling condition, residence prior to entry into the homeless program, zip code of the last residence, housing status, program entry date, program exit data, a unique person identifier, and household identification number. HUD revised the data standards in 2010 to reconcile them with provisions of the American Recovery and Reinvestment Act of 2009. A description of the current standards and changes made in response to ARRA can be found at http://www.hudhre.info/documents/FinalHMISDataStandards_March2010.pdf. For another good discussion of the data standards for the HMIS system, see http://www.lahsa.org/docs/hmis/HMIS%20Data%20Standards%20FAQ.pdf.

Note that programs funded by the Violence Against Women Act are prohibited from collecting certain identifying information. See the section entitled "HMIS Data and Preemption" below for more detail.

HMIS Information and Researcher Access

PPI can be disclosed externally or used internally by the homeless organization only if the use or disclosure is permitted by law and the use or disclosure is described in the organization's privacy policy. One of the permitted uses of PPI is for academic research (Privacy Standard 4.1.3). There must be a written research agreement between the HMIS provider and the researcher (again, similar conceptually to the data use agreements required by other federal laws). The research agreement must establish rules and limitations for processing and maintaining the security of PPI, provide for its return or disposal at the end of the research, restrict additional use

or disclosure, and require the data recipient to agree to abide by the conditions. An independent privacy consultant retained by HUD to work on HMIS issues has prepared a detailed discussion of HMIS research agreements, including a model agreement at http://www.hmis.info/ClassicAsp/documents/Research%20Disclosure%20Template%20-%20Gellman%201-4%20with%20watermark.pdf.

Alternatives to Use of PPI Under HMIS

As this brief overview suggests, HMIS data, including protected personal information (PPI), are available for research. However, in many cases researchers will wish to work with de-identified data from the HMIS system. HUD has published a paper providing guidelines on unduplicating and de-identifying records in the HMIS system (Sokol, 2005, available at http://www.hmis.info/ClassicAsp/documents/Technical%20Guidelines%20for%20Unduplicating%20and%20De-Identifying%20HMIS%20Client%20Records.pdf). Sokol provides a good discussion of the potential conflict between the goal of producing an unduplicated count of people who have used homeless services and privacy and confidentiality concerns, and provides good information about techniques to overcome these issues.

THE CHILD ABUSE PREVENTION AND TREATMENT ACT (CAPTA)

Federal law establishes strict confidentiality requirements for child abuse records. The Child Abuse Prevention and Treatment Act (CAPTA) is the core statute that establishes confidentiality rules that state programs must follow with child abuse records (see generally http://uscode.house.gov/download/pls/42C67.txt). Section 5106(b) sets out the confidentiality provisions, which are explained by the Administration of Children and Families in the U.S. Department of Health and Human Services at http://www.acf.hhs.gov/cwpm/programs/cb/laws_policies/laws/cwpm/questDetail.jsp?QAId=67. The Administration of Children and Families has also prepared a comprehensive guide to CAPTA that can be found at http://www.acf.hhs.gov/programs/cb/laws_policies/cblaws/capta03/capta_manual.pdf).

Child abuse records also may be subject to federal provisions governing the confidentiality of information on people applying for and/receiving federal assistance from social security and other federal programs (see 45 CFR 205.50 for the rule, at http://edocket.access.gpo.gov/cfr_2008/octqtr/45cfr205.50.htm).

CAPTA and Researcher Access to Child Abuse Records

CAPTA directs the Secretary of HHS to "carry out a continuing interdisciplinary program of research, including longitudinal research, that is designed to provide information needed to better protect children from abuse or neglect and to improve the well-being of abused or neglected children, with at least a portion of such research being field initiated" (42 USC 5105). CAPTA then defines in considerable detail the types of research the secretary is to stimulate. For those interested in using child abuse and child welfare records in research, there are various websites that provide good information about available data and how to access it. See, for example, the National Data Archive on Child Abuse and Neglect maintained by Cornell University at http://www.ndacan.cornell.edu/ and the website of Chapin Hall at the University of Chicago, which does extensive research using child welfare datasets, among other work: http://www.chapinhall.org/.

STATE LAW ISSUES; PREEMPTION

While federal law establishes confidentiality standards for many types of information, state law is the primary source of standards for other types. Some different types of records likely to be governed by state law are discussed below. In addition, there are occasions when the federal government has stated that federal law applies, absent special circumstances, that is, that federal law "preempts" state law. One example, discussed below, is found in HIPAA which says that HIPAA rules apply to the disclosure of PHI by a covered entity unless state confidentiality standards are more stringent. HIPAA is not the only law that permits states to adopt more stringent standards; FERPA does as well. In such cases, states must at least meet federal standards, though they may exceed them.

Criminal and Juvenile Justice Records

State laws typically govern access to criminal records, such as arrest records, and juvenile justice records, such as juvenile court files. All states traditionally have sought to protect the confidentiality of records that are generated by the juvenile justice system, and state law must be consulted to determine their availability. A collaboration between the United States Office of Juvenile Justice and Delinquency Prevention and Fox Valley Technical College provides links to each state's laws on juvenile interagency information and record sharing and can be found here http://dept.fvtc.edu/childprotectiontraining/states.htm. While adult arrest records are usually considered public records, states often have myriad rules on the confidentiality

of and access to court records and filings. Michigan is an example of a state that provides an excellent guide to its laws and court rules on this topic at http://www.courts.michigan.gov/scao/resources/standards/cf_chart.pdf.

Mental Health Records

All states have statutory provisions governing the confidentiality of mental health records. Treatment providers may be covered entities under HIPAA and if so, their treatment records will probably meet the definition of protected health information (PHI) established by HIPAA. This presents an illustration of the issue of preemption. If the state mental health confidentiality provisions are stronger than those in HIPAA, then the *state* law applies. At the same time, many state mental health laws permit unconsented disclosure for the purpose of research. The New York statute provides an example (N.Y. 33.13 (c) 9(iii)). It permits the Commissioner of the State Office of Mental Health to authorize release of information to:

> qualified researchers upon the approval of the institutional review board or other committee specially constituted for the approval of research projects at the facility, provided that the researcher shall in no event disclose information tending to identify a patient or client.

If the State Office of Mental Health is a covered entity and is permitting the release of PHI, then HIPAA provisions apply unless Michigan law provides for more stringent protections.

HIV Laws

Most states also have special laws protecting the confidentiality of information that may disclose a person's HIV status. The underlying assumption is that disclosure could lead to discrimination and stigmatization of the affected person (Doughty, 1994). But as is also the case with mental health laws, HIV confidentiality laws may permit disclosure of personally identifying information to researchers in some circumstances. For example, California Health and Safety Code Section 121025(b), addressing the disclosure of HIV information, provides:

> In accordance with subdivision (f) of Section 121022, a state or local public health agency, or an agent of that agency, may disclose personally identifying information in public health records, as described in subdivision (a), to other local, state, or federal public health agencies or to corroborating medical researchers, when the

> confidential information is necessary to carry out the duties of the agency or researcher in the investigation, control, or surveillance of disease, as determined by the state or local public health agency.

In determining whether HIPAA or California law applies in this case, the starting point would be whether the "state or local public health agency" is a covered entity. If so, then HIPAA would apply if its confidentiality protections are more stringent than this provision of state law. However, if the public health agency is *not* a covered entity, then state law would apply, as HIPAA would not be applicable.

HMIS Data and Preemption

When the HMIS system was mandated, advocates and communities were concerned that reporting some required information, particularly the identifying residence, would endanger victims of domestic abuse. As a result, HUD specified that if a state law provides stronger confidentiality protections for domestic violence service providers, then the state law prevails. In the Violence Against Women Act (VAWA) of 2005, Congress specifically prohibited the disclosure of *personal information* by programs receiving VAWA grants absent informed consent by the individual (42 USC §11383(a)(18)). Personal information includes individually identifying information for or about an individual, including information likely to disclose the location of a victim of domestic violence, dating violence, sexual assault, or stalking, including:

a. a first and last name;

b. a home or other physical address;

c. contact information (including a postal, e-mail or Internet protocol address, or telephone or facsimile number);

d. a social security number; and

e. any other information, including date of birth, racial or ethnic background, or religious affiliation, that, in combination with any of subparagraphs (A) through (D),

d. would serve to identify any individual.

Therefore in this case, the federal provision now prevails, and states are prohibited (that is, preempted) from requiring the disclosure of this information.

Medicaid Records

Access to Medicaid records is generally controlled by the state agency that administers the Medicaid program. If the state Medicaid agency is a covered entity, then HIPAA comes into play, and the application of state or federal law depends on which provides the most stringent confidentiality protections, as discussed above. There is great variability among states in conditioning access to Medicaid data, though the majority of states do provide access in some fashion. For an excellent review see Stiles, Boothroyd, Robst, and Ray, 2011, at http://aas.sagepub.com/content/43/2/171.full.pdf+html.

Research Approval: Institutional Review Boards (IRBs) and Privacy Boards (HIPAA)

INSTITUTIONAL REVIEW BOARDS (IRBs)

All researchers are familiar with Institutional Review Boards (IRBs). The Common Rule defines their mission as the protection of the rights, welfare, and privacy of research subjects. While the discussion below addresses current IRB requirements for studies using individual information and other forms of data, HHS has proposed a number of changes to the Common Rule, including specifying data security protections tied to the level of identifiability of the data and altering IRB review rules for some types of social studies. A table comparing the current rules with the proposed rules can be found here http://www.hhs.gov/ohrp/humansubjects/anprmchangetable.html. What is of particular interest is that the Common Rule would incorporate the levels of data established in HIPAA (individually identifiable, limited dataset, and de-identified data). All levels of studies would have to have data security protections commensurate with the level of data used in the study, and IRBs would be relieved of the burden of assessing the informational risk presented by each study.

At present, the IRB must assure that "when appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data" (46.111(a) (7)). A "human subject" is defined as "a living individual about whom an investigator … conducting research obtains (1) data through intervention or interaction with the individual, or (2) identifiable private information" (45 CFR §46.102(f)). The level of IRB scrutiny varies depending on the type of research and the level of risk to the human subject. Research that presents more than minimal risk

to the subject will require full IRB review; minimal risk means that the probability and magnitude of harm or discomfort anticipated in the research are not greater, in and of themselves, than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations (46.102(i)). Some research is exempt from IRB review, some qualifies for expedited review, and some requires full review. Examples follow. In addition, for a helpful, chart-based analysis of whether IRB approval is required, and if so, whether the proposal is eligible for expedited or full review, see this HHS link: http://www.hhs.gov/ohrp/policy/checklists/decisioncharts.html#c3.

Exempt Research

Research that involves normal educational practices that will occur *only* in commonly established or commonly accepted educational institutions is exempt from IRB review (45 CFR 46.101(b) (1). Research involving the use of educational tests, survey procedures, interview procedures, or observation of public behavior is also exempt (45 CFR 46.101(b)(2). And (b)(3)) if the investigator does not record the information in a manner that permits the identification of human subjects directly or through identifiers linked to the subjects and any disclosure of a subject's responses outside the research could not reasonably place the subject at risk of liability or damage their financial standing, employability, or reputation (a proviso that applies as well to the following exempt research). Other exempt research includes studies using only the collection or study of existing data, including documents or records (45 CFR 46.101(b) (4)), and research or demonstration projects conducted or approved by one of the Common Rule agency heads involving public benefit or service programs, procedures for obtaining benefits under those programs, or changes in or alternatives to the programs or methods or levels of payment (45 CFR 46.101(b) (5)).

Researchers may wish to use data contained in clinical registries and warehouses for their studies. It might be argued that such registries involve "existing data" and so should be exempt from IRB review. However, Dokholyan and colleagues argue that the use of clinical registries requires IRB review because they contain identifying links necessary to link regularly updated data (Dokholyan, Muhlbaier, Faletta, et al, 2009). This article provides a very thorough and interesting overview of the issues involved in linking clinical and administrative databases.

Expedited Review

If a study qualifies for expedited review, the Chair of the IRB or another member designated by the Chair may review the study in lieu of the full IRB. To qualify, a study must present minimal risk to the subjects and be one of those listed in the regulation (45 CFR 46.110) as eligible for expedited review. Note that expedited review may not be used where identification of the subjects or their responses would place them at risk for liability or damage to their reputation, employability, insurability, or be stigmatizing unless the investigator takes steps to assure that the risk is no greater than minimal. Most of the categories eligible for expedited review will be of little interest to researchers using linked datasets, for example, some clinical studies, collection of blood samples and biological specimens, and collection of data through non-invasive medical procedures. However, other categories may be of interest. These include expedited review for research involving materials such as data, documents, or records collected only for non-research purposes; collection of data from voice, video, digital, or image recordings made for research purposes; and research on individual or group characteristics or behavior or research using data collection techniques including "program evaluation" or "quality assurance" methodologies. A "guidance on expedited review procedures" can be found on the HHS website here http://www.hhs.gov/ohrp/policy/exprev.html.

Full Review

Studies not qualifying for exempt or expedited status require full review by the IRB. Studies involving "vulnerable populations" (pregnant women, prisoners, children) will almost always require full review. A risk of breach of privacy (relevant to studies relying on existing or new datasets with identifiers) may constitute more than minimal risk, resulting in full IRB review of the proposed study.

PRIVACY BOARDS

HIPAA has created additional requirements for studies involving PHI. The Privacy Rule requires that the IRB must determine that, when appropriate, the research protocol includes "adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data" (see 45 CFR 46.111(a)(7) and 21 CFR 56.111(a)(7)). The IRB can assume the responsibilities of a Privacy Board. Privacy Boards acting in lieu of IRBs in addressing privacy issues in research have similar authority and obligations. The National Institutes of Health have prepared fact sheets on HIPAA and Privacy Boards.

Visit (http://privacyruleandresearch.nih.gov/privacy_boards_hipaa_privacy_rule.asp) and IRBs (http://privacyruleandresearch.nih.gov/irbandprivacyrule.asp).

Not all researchers are affected by the Privacy Rule, nor is all research. If a researcher works directly for a covered entity, then the Privacy Rule applies, including the Privacy Board provisions (again, note that an IRB can act as a Privacy Board). If a researcher is *not* working for a covered entity, but plans on using PHI, then the researcher will be affected by the Privacy Rule because the covered entity will not be able to release PHI to the researcher unless the conditions noted below are met.

Consent/Waiver of Consent (IRBs); Authorization/Waiver of Authorization (PHI)

Generally, a research subject must provide consent to participation in a study. The Office for Human Research Protections (OHRP) in the U.S. Department of Health and Human Services provides an abundance of information about consent in research at http://www.hhs.gov/ohrp/policy/consent/index.html. In addition, the Agency for Healthcare Research and Quality (AHRQ) has prepared a thorough and useful guide titled *Informed Consent and Authorization Toolkit for Minimal Risk Research* that can be found here: http://www.ahrq.gov/fund/informedconsent/ictoolkit.pdf

If protected health information (PHI) is involved, then the subject must provide authorization. HIPAA has specific requirements for authorizations. Note that other regulations (such as 42 CFR Part 2, governing alcohol and drug records, state health and mental health confidentiality laws, HIV confidentiality laws, etc.) also have specific elements for disclosure that consent forms must contain. An example of a federal form used by the Social Security Administration that permits disclosure of PHI, substance use/alcohol records and education records can be found at http://www.ssa.gov/online/ssa-827.pdf.

Obtaining individual consent can become burdensome when linked datasets are being used in research, given the impracticalities of obtaining consent from the thousands of people whose information may be stored. The IRB may waive consent in two circumstances. One involves the study of certain aspects of public benefit programs (45 CFR 46.116(c1).

The IRB may also waive, or alter consent procedures when:

1. The research involves no more than minimal risk to the subjects;
2. The waiver or alteration will not adversely affect the rights and welfare of the subjects;

3. The research could not practicably be carried out without the waiver or alteration; and
4. Whenever appropriate, the subjects will be provided with additional pertinent information after participation (45 CFR 46.116(c) (2).

In seeking a waiver or alteration, the investigator using linked datasets will have to illustrate how data are protected from intrusion as part of showing that the study involves no more than minimal risk.

If protected health information (PHI) and HIPAA are involved, the researcher will need to obtain a waiver of authorization in lieu of a waiver of consent. If the researcher intends to use PHI but it is impracticable to obtain individual authorization from those whose PHI is being utilized, the IRB has the authority to waive or alter the Privacy Rule's authorization requirements, in whole or in part. The IRB may waive the authorization requirement when it finds the following (see 164.512(i), as well as the discussion by the National Institutes of Health at http://privacyruleandresearch.nih.gov/irbandprivacyrule.asp):

a. The use or disclosure of protected health information involves no more than minimal risk to the individuals;
b. The alteration or waiver will not adversely affect the privacy rights and the welfare of the individuals;
c. The research could not practicably be conducted without the alteration or waiver;
d. The research could not practicably be conducted without access to and use of the protected health information;
e. The privacy risks to individuals whose protected health information is to be used or disclosed are reasonable in relation to the anticipated benefits, if any, to the individuals, and the importance of the knowledge that may reasonably be expected to result from the research;
f. There is an adequate plan to protect the identifiers from improper use and disclosure;
g. There is an adequate plan to destroy the identifiers at the earliest opportunity consistent with conduct of the research, unless there is a health or research justification for retaining the identifiers, or such retention is otherwise required by law; and
h. There are adequate written assurances that the protected health information will not be reused or disclosed to any other person

> or entity, except as required by law, for authorized oversight of the research project, or for other research for which the use or disclosure of protected health information would be permitted by this subpart.

Many IRBs have created forms for researchers to use in applying for a waiver of authorization when PHI is involved. For example, the Veteran's Administration form can be found here: http://hipaa.wisc.edu/ResearchGuide/forms/WaiverAuthAppVA_fill.pdf

A similar form, used by the University of Connecticut Health Center can be found here: http://www.policies.uchc.edu/policies/hipaa_waiver_app_authorization.pdf. A sample form prepared by the California Pacific Medical Center permits a researcher to apply for one or both waivers, as appropriate, on the same form. See http://www.cpmc.org/professionals/research/irb/forms/form9-waiverofauthorization.html. See also a form available from the Drexel IRB that enables the investigator to seek a waiver of the Common Rule consent requirement and/or HIPAA's authorization requirement: http://www.research.drexel.edu/compliance/IRB/HIPAA.aspx.

A covered entity may not use or release PHI to a researcher until it receives documentation of:

- The identity of the approving IRB;
- The date on which the waiver or alteration was approved;
- A statement that the IRB has determined that all the specified criteria for a waiver or an alteration were met;
- A brief description of the PHI for which use or access has been determined by the IRB to be necessary in connection with the specific research activity;
- A statement that the waiver or alteration was reviewed and approved under either normal or expedited review procedures;
- The required signature of the IRB chair or the chair's designee.

The Privacy Rule does not require authorization or IRB approval of a waiver or alteration of authorization if the researcher is using a limited dataset (defined above). However, if the activity that relies on the limited dataset meets the definition of "research" then IRB approval under ordinary IRB rules is necessary; the IRB does not need to perform the extra activities established by the Privacy Rule. In addition, the researcher and covered entity must enter into a Data Use Agreement

as a predicate to disclosing a limited dataset, even when the researcher works for the covered entity. A Data Use Agreement establishes the ways in which data in the dataset will be used and protected. There are many examples online. One that also contemplates the use of identifiable information is that used by the Centers for Medicare and Medicaid Services: https://www.cms.gov/cmsforms/downloads/cms-r-0235.pdf. Another example is that used by the North Carolina Department of Health and Human Services: http://info.dhhs.state.nc.us/olm/manuals/dhs/pol80/man/DHHS_Data_Use_Agreement_Template.pdf.

Many have argued that HIPAA has had a negative impact on research, by making the IRB process more complicated and drawn out, and by making greater numbers of people unwilling to sign authorizations permitting their information to be used in research (Dunlop, Graham, Leroy, et al, 2007; Beebe, Ziegenfuss, Sauver, et al., 2011; Shalowitz, & Wendler, 2006; for a lengthy critique see IOM, 2009). Regardless of the truth of this argument, HIPAA has become a fact of life and as Stiles and Boothroyd point out in their paper for this series, researchers have looked for ways to accommodate their research to this new reality. In addition, the proposed amendments to the Common Rule, which would conform Common Rule and HIPAA data definitions and protections, would potentially ameliorate some of these issues.

The authorization form template used by the University of Pennsylvania for research purposes can be found at http://www.med.upenn.edu/ohr/docs/HIPAAauth.doc.

Data Management and Security; the HIPAA Security Rule

The management and security of data is an essential responsibility of researchers. It is a responsibility that perhaps looms even larger when using linked datasets with information on many individuals, whether identifiable or not.

At present, IRBs are not responsible for assuring data security maintained or accessed by researchers, nor do they have the technical expertise to do so. The proposed amendments to the Common Rule would attempt to protect research subjects from "information risk," discussed in the proposal in the following terms: Informational risks derive from inappropriate use or disclosure of information, which could be harmful to the study subjects or groups. For instance, disclosure of illegal behavior, substance abuse, or chronic illness might jeopardize current or future employment, or cause emotional or social harm. In general, informational risks are correlated with

the nature of the information and the degree of identifiability of the information. The majority of unauthorized disclosures of identifiable health information from investigators occur due to inadequate data security. http://www.gpo.gov/fdsys/pkg/FR-2011-07-26/html/2011-18792.htm.

HHS has proposed three specific strategies for tightening data security. The first would be to require research collecting identifiable data, as well as data in limited dataset form, to conform to security standards modeled on the HIPAA Security Rule (discussed immediately below). The second would permit researchers to view individual identifiers for limited datasets or de-identified data as long as the researcher does not record them; this currently is not permitted. Third, HHS would assure that periodic audits of data security and management occur.

Pending these changes, there are resources for considering strategies for data management and security. One useful site is maintained by the Office of Research Integrity of the U.S. Department of HHS; the site offers information and on-line materials regarding these issues. An example can be found here: http://ori.hhs.gov/education/products/clinicaltools/data.pdf.

Given that HHS proposes relying on the HIPAA Security Rule for guidance in guarding against "information risk," it is worth examining some features of the Security Rule. The Security Rule (45 CFR 164.306, found at http://www.hhs.gov/ocr/privacy/hipaa/administrative/securityrule/index.html) like the Privacy Rule, applies only to covered entities maintaining PHI in electronic form. The threshold question then is whether the researcher is independently a covered entity or works for one. The Security Rule provides:

> Researchers who are members of a covered entity's work force may be covered by the Security standards as part of the covered entity. See the definition of "workforce" at 45 CFR 160.103. Note, however, that a covered entity could, under appropriate circumstances, exclude a researcher or research division from its health care component or components (see § 164.105(a)). Researchers who are not part of the covered entity's workforce and are not themselves covered entities are not subject to the standards.

Some institutions, for example, the University of California, have declared themselves hybrid entities. At the University of California, those parts of the University providing health care are covered entities, but education and research are excluded and are not considered part of the "covered entity" of the University.

Therefore, a member of the University faculty who is *not* part of the health care functions within the University is not bound by the Security Rule, because he or she is not part of a covered entity. However, an employee of the University health care system conducting research would be part of the covered entity and the Security Rule would apply to PHI held by that researcher electronically. (For an explanation, see http://www.universityofcalifornia.edu/hipaa/docs/research_guidelines.pdf). OCR provides an explanation of a "hybrid entity" here: http://www.hhs.gov/ocr/privacy/hipaa/faq/research_disclosures/315.html.

If the linked datasets used by researchers do not contain PHI, then the Security Rule would not apply. If the Security Rule does apply, according to the United States Department of Health and Human Services, a covered entity has several core obligations (for a discussion, see (http://www.hhs.gov/ocr/privacy/hipaa/understanding/srsummary.html).

These obligations are to:

1. Ensure the confidentiality, integrity, and availability of all e-PHI they create, receive, maintain or transmit;
2. Identify and protect against reasonably anticipated threats to the security or integrity of the information;
3. Protect against reasonably anticipated, impermissible uses or disclosures; and
4. Ensure compliance by their workforce.

The Security Rule has three general domains, each one containing various rules. These domains include *administrative safeguards*, *physical safeguards*, and *technical safeguards*. Within each domain, there are "implementation specifications." Some of these are required, while others are "addressable". Covered entities must comply with the former, but have discretion in complying with the latter. However, before exercising that discretion, the covered entity must perform an analysis to determine whether the implementation specification is a reasonable and appropriate safeguard for implementation in the covered entity's environment. After the analysis is performed, the covered entity may implement the specification, implement an alternative, or do neither. In any case, the analysis and decision must be documented.

ADMINISTRATIVE SAFEGUARDS

There are nine standards within the Administrative Safeguards section of the Security Rule. The standards cover a variety of topics, focusing on workforce issues, risk analysis, and contingency planning, among others. The most important required specifications are to conduct a risk analysis of security issues on an on-going basis (164.308(a)(1)); to assign responsibility for issues to a particular person (164.308(a)(3)); and to have contingency planning for emergencies (164.308(a)(7)). The risk analysis, which is at the heart of the Security Rule, must be "an accurate and thorough assessment of the potential risks and vulnerabilities to the confidentiality, integrity, and availability of electronic protected health information held by the covered entity." Whether an electronic data system has PHI or not, it is worth considering having a third party conduct a risk analysis periodically to ensure that the system continues to be adequately protected against technological and other intrusions.

PHYSICAL SAFEGUARDS

There are four standards in this section of the rule, which addresses workstation issues and controlled access to data. The covered entity is required to address workstation use and security, as well as the disposal and re-use of data (164.308(b), (c), and (d)(1)).

TECHNICAL SAFEGUARDS

There are five standards in this section, addressing access control, audit controls, data integrity, person or entity authentication, and transmission security (164.312(a)-(e).

The Department of Health and Human Services has provided a good introductory comparison of the Privacy and Security Rules at http://www.hhs.gov/ocr/privacy/hipaa/administrative/securityrule/security101.pdf. In addition, The National Institute of Standards and Technology has published an essential resource guide to the Security Rule, with detailed analysis and suggestions for implementing the rule which can be found here: http://csrc.nist.gov/publications/nistpubs/800-66-Rev1/SP-800-66-Revision1.pdf.

While only covered entities are bound by the Security Rule, those not bound by it may wish to follow the evolution of Health and Human Service's proposed amendments to the Security Rule, since as discussed above, those amendments rely to some degree on the Security Rule's standards for data security.

Penalties and Enforcement

There are several ways that laws can be enforced. These include civil lawsuits for damages brought by individuals whose rights were violated, enforcement actions brought by administrative agencies charged with enforcement of a particular law, and in some cases criminal actions. This is true of confidentiality and privacy laws as well.

PRIVATE CAUSES OF ACTION

Typically, individuals can bring lawsuits for monetary damages for breach of their rights only if the statute creating the right permits such lawsuits. In other words, if a person's HIPAA rights are violated, that person can bring a lawsuit only if HIPAA permits an individual to do so. Many of the statutes discussed in this paper do not create a private right to sue and therefore individual lawsuits cannot proceed. For example, courts have ruled that an individual does not have a private cause of action (that is, the right to bring a lawsuit for personal damages) under HIPAA (Acara v. Banks, 470 F. 3rd 569, 2006, provides one example). The same is true of FERPA (Gonzaga v. Doe, 563 U.S. 273, 2002); and 42 CFR Part 2 (Chapa v. Adams, 168 F. 3rd 1036). The federal Privacy Act of 1974 permits some types of damage claims by private citizens. The United States Department of Justice maintains a website discussing the Privacy Act, including enforcement provisions such as civil remedies, at http://www.justice.gov/opcl/1974privacyact-overview.htm. Whether an individual can sue for breach of confidentiality or privacy rights created by *state* law will depend on whether the state law in question creates such rights. For example, the California Supreme Court ruled recently that an individual could sue a debt collector for disclosing his and his children's dental records to credit reporting agencies under the state's Confidentiality of Medical Information Act (Brown v. Mortensen, S180862, 2011, at http://www.courtinfo.ca.gov/opinions/documents/S180862.PDF).

There also have been lawsuits alleging damages resulting from breaches of data affecting classes (or groups) of individuals. One example is Pisciotta v. Old Nat'l Bancorp (439 F. 3rd 629, 2007) in which the United States Court of Appeals for the 7th Circuit dismissed a lawsuit brought by a class of individuals whose personal bank data had been stolen. The Court said that no injury had resulted from illegal use of the data so the case could not proceed. On the other hand, the 9th Circuit Court of Appeals ruled that a lawsuit could proceed in the absence of actual injury in a case

stemming from the theft of a laptop computer containing personal information about Starbucks employees (Krottner v. Starbucks, 628 F.3d 1139, 2010). While the courts are split on whether actual harm must be alleged for a class action case to proceed, one can anticipate that such lawsuits will continue to be filed after large data breaches.

ADMINISTRATIVE ENFORCEMENT

Different agencies are charged with primary enforcement of the federal statutes and regulations discussed in the paper. The United States Department of Justice is primarily responsible for enforcement of the Privacy Act of 1974. The Office for Civil Rights in the Department of Health and Human Services is primarily responsible for enforcement of HIPAA (see http://www.hhs.gov/ocr/privacy/hipaa/enforcement/index.html) though United States Attorneys may enforce HIPAA, and the Health Information Technology for Economic and Clinical Health Act of 2009 (HITECH) expanded enforcement authority to state Attorneys Generals. The Family Compliance Office (FPCO) of the United States Department of Education has primary responsibility for enforcing FERPA (http://www2.ed.gov/policy/gen/guid/fpco/index.html). Reports of violations of 42 CFR Part 2 may be made to the local United States Attorney's office (42 CFR 2.5). Primary enforcement of state laws is determined by state law.

PENALTIES: CIVIL AND CRIMINAL BROUGHT BY ENFORCEMENT AGENCIES

Enforcing agencies can seek civil or criminal penalties, depending on the penalty scheme established in statute and/or regulation, the seriousness of the alleged breach, and the discretion of the agency. HIPAA is used as an example here because of recent strengthening of its penalty provisions in the Health Information Technology for Economic and Clinical Health (HITECH) Act. The penalty schemes for violations of other statutes are available at the websites of the agencies charged with primary enforcement.

HITECH significantly increased both per violation and aggregate **civil** penalties for HIPAA violations. HIPAA originally provided civil fines of $100 per violation, with the total for all violations of an identical requirement not to exceed $25,000 in a year. The new provisions now provide for categories of violations, with increasing penalties tied to intent of the violating party and harm to the party whose interests were violated. The maximum penalty per violation is now $50,000 with a total

annual cap of $1.5 million. The HIPAA Privacy and Security Rules may now be directly enforced against Business Associates. If a party is aware of violations and does not take steps to correct them, then the penalties will be at the top of the range; if steps are taken within thirty days to correct violations caused by willful neglect, the penalties (depending on harm suffered by the injured party) are more likely to be in the lower part of the range ($10,000 to $50,000).

In a sign that the Obama administration intends to enforce HIPAA more vigorously in at least some cases, the Office for Civil Rights announced a $4.3 million fine against Cignet Health in Maryland, after Cignet had refused to produce records for forty-one patients who had requested those records, and also refused to cooperate with the federal investigation (http://www.hhs.gov/news/press/2011pres/02/20110222a.html).

Criminal penalties are available for *knowing* violations of the law. If a person knowingly discloses PHI to another person, with intent to sell, transfer, or use PHI for commercial advantage, personal gain, or malicious harm, this is a crime punishable by up to ten years of imprisonment and fines up to $250,000 (42 USC 1320d-6[b][3] [2010]). For other disclosures that the disclosing party knew were illegal, possible criminal penalties include fines up to $50,000 and up to one year imprisonment (42 USC 1320d-6[b][1][2010]).

There have been few criminal cases brought under HIPAA but those that exist are illustrative of the type of conduct that investigators are likely to examine. In a recent case, a UCLA physician pled guilty to inappropriately accessing medical records and will serve four months in prison. According to accounts of the case, he looked at patient records inappropriately 323 times in one week. In an earlier case, an individual pled guilty to obtaining records of a patient from his employer and then used information in the record to obtain credit cards, eventually incurring more than $9,000 in debt (http://www.crowell.com/NewsEvents/Newsletter.aspx?id=546). In a third case, a health care employee accessed a patient's records and provided information to (the employee's) husband for the husband's use in a private lawsuit against the patient. The employee pled guilty and received two years probation and community service (http://www.kuar.org/kuarnews/4433-criminal-enforcement-of-hipaa-still-a-new-concept.html). In two of these cases, personal gain was involved; in the other, there was clearly a pattern of violating patients' confidentiality and privacy rights.

The Office for Civil Rights provides ongoing updates on its enforcement activities at http://www.hhs.gov/ocr/privacy/hipaa/enforcement/highlights/index.html. According

to OCR, through February 2011, it had received 58,911 complaints alleging HIPAA violations. More than 33,000 of these cases were not eligible for enforcement. In another 6,784 cases, OCR investigated and found no violation. Finally, OCR had performed some type of enforcement action in 13,003 cases. However, enforcement often took the form of an advisory letter to the covered entity.

The most common complaints in order of frequency have been:

1. Impermissible uses and disclosures of protected health information;
2. Lack of safeguards of protected health information;
3. Lack of patient access to their protected health information;
4. Uses or disclosures of more than the minimum necessary protected health information; and
5. Complaints to the covered entity that were ignored.

Summary

Electronic data systems linking multiple datasets are a rich source of information for many different types of researchers. These datasets may contain myriad types of information, some that identify an individual, and some which do not. Various laws, both federal and state, are relevant to the confidentiality, privacy, and disclosure of such information. Laws for discrete types of information may use different definitions, create different rules for disclosure, and are sometimes difficult to understand. However, none of these laws create a categorical prohibition against using sensitive identifying information for research purposes (though access may be limited) and several provide explicit guidance on de-identifying data. Information can be retrieved and retained, but it must be done so securely. HIPAA has detailed rules for the security of electronic PHI, but even if a data system does not contain PHI, or principally holds de-identified data, the security of the system must be addressed. In fact, maintaining the security of electronic data is the best safeguard for assuring its continuing confidentiality once it has been made available for research. This is why the researcher must be cognizant not only of the rules regarding confidentiality, but assure security commensurate with the risk posed if data is inappropriately accessed, if data integrity is compromised, or if data are misused or improperly disclosed. Privacy and security are essential factors in managing data, both for researchers and those holding data that researchers may wish to access.

Appendix: Useful Websites

The Common Rule (Governing Conduct of Research in Most Situations)

Discussion of the Common Rule:
http://www.hhs.gov/ohrp/humansubjects/commonrule/index.html.

Discussion of proposed amendments to the Common Rule by the Department of Health and Human Services: http://www.hhs.gov/ohrp/humansubjects/anprm2011page.html.

Privacy Act of 1974

U.S. Department of Education requirements for Privacy Act matching agreements:
http://www2.ed.gov/policy/gen/leg/foia/acsom6105.pdf

Access to Medicare data under the Privacy Act:
http://www.resdac.org/medicare/medicarefaq.asp

HIPAA (Protected Health Information)

HHS guidance to covered entities: http://www.resdac.org/medicare/medicarefaq.asp

HHS/Education guidance to relationship between FERPA and HIPAA:
http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/hipaaferpajointguide.pdf.

National Institutes of Health discussion of clinical research and Privacy Rule:
http://privacyruleandresearch.nih.gov/pr_02.asp

Office for Civil Rights discussion of HIPAA and research:
http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/research.html.

HHS discussion of de-identification of health information:
http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/deidentificationworkshop2010.html)

Centers for Medicare and Medicaid Services data use agreement:
https://www.cms.gov/cmsforms/downloads/cms-r-0235.pdf

North Carolina Department of Health and Human Services data use agreement:
http://info.dhhs.state.nc.us/olm/manuals/dhs/pol-80/man/DHHS_Data_Use_Agreement_Template.pdf

Data use agreement example for a limited dataset: http://info.dhhs.state.nc.us/olm/manuals/dhs/pol-80/man/DHHS_Data_Use_Agreement_Template.pdf

University of Buffalo explanation for why business associate agreements are not required for researchers: http://www.hpitp.buffalo.edu/hipaa/Research/DataExtraction.htm

HHS discussion of business bssociates, noting that researchers are not required to enter business associate agreements for the purpose of accessing protected health information for research: http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/businessassociates.html, Other Situations in Which a Business Associate Contract is NOT Required

FERPA (Education Records)

United States Department of Education guidance on FERPA and resources: http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html

HHS/Department of Education guidance to relationship between FERPA and HIPAA: http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/hipaaferpajointguide.pdf

United States Department of Education guidance on protection of human subjects: http://www2.ed.gov/about/offices/list/ocfo/humansub.html

U.S. Department of Education sample agreement between educational institution and authorized representative: http://www2.ed.gov/about/offices/list/ovae/pi/cte/uiferpa.html.

Amended FERPA regulation permitting data sharing agreements with entities not under "direct control" of the educational institution: http://www.gpo.gov/fdsys/pkg/FR-2011-12-02/pdf/2011-30683.pdf

National Center for Educational Statistics guide to privacy and confidentiality of educational records: http://nces.ed.gov/pubs2011/2011601.pdf

42 CFR (Substance Abuse and Alcohol Records)

Discussion of the relationship between the HIPAA Privacy Rule and 42 CFR: http://www.gwumc.edu/sphhs/departments/healthpolicy/CHPR/downloads/behavioral_health/bhib-18-19.pdf

HMIS (homeless management information system)

Overview of the Homeless Management Information System (HMIS) prepared by the United States Department of Housing and Urban Development: http://portal.hud.gov/hudportal/HUD?src=/program_offices/comm_planning/homeless/hmis).

Overview of data elements that must be collected by HMIS programs: http://www.hudhre.info/documents/FinalHMISDataStandards_March2010.pdf

Discussion of data standards for HMIS system: http://www.lahsa.org/docs/hmis/HMIS%20Data%20Standards%20FAQ.pdf

Discussion with example of HMIS research agreements: http://www.hmis.info/ClassicAsp/documents/Research%20Disclosure%20Template%20-%20Gellman%201-4%20with%20watermark.pdf

Discussion of deidentified protected personal information (PPI) in the HMIS system: http://www.hmis.info/ClassicAsp/documents/Technical%20Guidelines%20for%20Unduplicating%20and%20De-Identifying%20HMIS%20Client%20Records.pdf).

Child Abuse Prevention and Treatment Act (CAPTA)

Confidentiality provisions of CAPTA: http://www.acf.hhs.gov/cwpm/programs/cb/laws_policies/laws/cwpm/questDetail.jsp?QAId=67

Guide to CAPTA prepared by the Administration of Children and Families of the United States Department of Health and Human Services: http://www.acf.hhs.gov/programs/cb/laws_policies/cblaws/capta03/capta_manual.pdf

Discussion of the use of child abuse records in research: http://www.ndacan.cornell.edu

Discussion of the use of child welfare records in research: http://www.chapinhall.org/

Criminal and Juvenile Justice Records

State laws on juvenile interagency information sharing:
http://dept.fvtc.edu/childprotectiontraining/states.htm

Guide to Michigan law and court rules on accessing court records and filings: http://www.courts.michigan.gov/scao/resources/standards/cf_chart.pdf

Institutional Review Boards (IRBs)

Comparison of current Common Rule regulatory provisions with changes to the Common Rule proposed by HHS: http://www.hhs.gov/ohrp/humansubjects/anprmchangetable.html

Overview of requirements on full or expedited review of studies by IRBs:
http://www.hhs.gov/ohrp/policy/checklists/decisioncharts.html#c3

Guidance on expedited review procedures: http://www.hhs.gov/ohrp/policy/exprev.html.

Fact sheets on HIPAA and Privacy Boards:
http://privacyruleandresearch.nih.gov/privacy_boards_hipaa_privacy_rule.asp

Fact sheets on IRBs and Privacy Rule:
http://privacyruleandresearch.nih.gov/irbandprivacyrule.asp

Consent/Authorization to research participation

HHS Office for Human Research Protections guidance on consent to research:
http://www.hhs.gov/ohrp/policy/consent/index.html

University of Pennsylvania form for authorization of release of protected health information (PHI): http://www.med.upenn.edu/ohr/docs/HIPAAauth.doc

Agency for Healthcare Research and Quality Informed consent and authorization toolkit for minimal risk research: http://www.ahrq.gov/fund/informedconsent/ictoolkit.pdf

Social Security Administration consent form covering protected health information, substance use/alcohol records, and educational records: http://www.ssa.gov/online/ssa-827.pdf

Veteran's Administration form for waiver of authorization in seeking data for research:
http://hipaa.wisc.edu/ResearchGuide/forms/WaiverAuthAppVA_fill.pdf.

University of Connecticut Health Center form for waiver of authorization:
http://www.policies.uchc.edu/policies/hipaa_waiver_app_authorization.pdf

California Pacific Medical Center form for waiver of authorization and/or consent:
http://www.cpmc.org/professionals/research/irb/forms/form9-waiverofauthorization.html.

Drexel University form for waiver of authorization and/or consent:
http://www.research.drexel.edu/compliance/IRB/HIPAA.aspx

Data security

The HIPAA Security Rule: http://www.hhs.gov/ocr/privacy/hipaa/administrative/securityrule/index.html

Discussion and materials for managing data security prepared by the Office of Research Integrity of HHS: http://ori.hhs.gov/education/products/clinicaltools/data.pdf

HHS discussion of "hybrid entities," that is, entities that have a component that is a covered entity and another that is not:
http://www.hhs.gov/ocr/privacy/hipaa/faq/research_disclosures/315.html

HHS discussion of the obligations of covered entities under Security Rule:
http://www.hhs.gov/ocr/privacy/hipaa/understanding/srsummary.html

HHS comparison of the HIPAA Privacy and Security Rules:
http://www.hhs.gov/ocr/privacy/hipaa/administrative/securityrule/security101.pdf

National Institute of Standards and Technology resource guide to the Security Rule can be found here: http://csrc.nist.gov/publications/nistpubs/800-66-Rev1/SP-800-66-Revision1.pdf

Enforcement of confidentiality laws

The HHS Office of Civil Rights is primarily responsible for enforcing HIPAA. It maintains a website on its enforcement activities here:
http://www.hhs.gov/ocr/privacy/hipaa/enforcement/index.html

The U.S. Department of Education's Family Compliance Office has primary responsibility for enforcing FERPA violations. Its website is here:
http://www2.ed.gov/policy/gen/guid/fpco/index.htm

References

Beebe, T.J., Ziegenfuss, J.Y., Savuer, J.L., Jenkins, S.M., Haas, L., Davern, M.E., & Talley, N.J. (2011). Health Insurance Portability and Accountability Act (HIPAA) and Survey Nonresponse Bias. *Medical Care*, 49, 365-370).

Benitez, K., & Malin, B. (2010). Evaluating re-identification risks with respect to the HIPAA privacy rule. *Journal of the American Medical Informatics Association*, 17 (2), 169-177.

Department of Health and Human Services (2005). Federal Committee on Statistical Methodology. *Statistical Policy Working Paper* 22. At http://www.fcsm.gov/workingpapers/spwp22.html

Dokholyan, R., Muhlbaier, L., Faletta, J., Jacobs, J., Haan, C., & Peterson, E. (2009). Regulatory and ethical considerations in linking clinical and administrative databases. *American Heart Journal*, 157, 971-982.

Doughty, R. (1994). The confidentiality of HIV-related information: Responding to the resurgence of aggressive public health interventions in the AIDS epidemic. *California Law Review*, 82, 111-184.

Dunlop, A.L., Graham, T., Leroy, Z., Glanz, K., & Dunlop, B. (2007). The impact of HIPAA authorization on willingness to participate in clinical research. *Ann Epidemiol 2007*;17:899–905.

Francis, L.P. (2008). Privacy and confidentiality: The importance of context. *The Monist*, 91 (1), 52-67.

Institute of Medicine of the National Academies (2009). *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. Washington DC.

Scholl, M., Stine, K., Hash, J., Bowen, P., Johnson, A., Smith, C.D., & Steinberg, D. (2008) *An Introductory Resource Guide for Implementing the Health Insurance Portability and 52 Accountability Act (HIPAA) Security Rule*. Computer Security Division United States Department of Commerce. Washington, DC: http://csrc.nist.gov/publications/nistpubs/800-66-Rev1/SP-800-66-Revision1.pdf

Schuurman, N., Morad Hameed, S., Fiedler, R., Bell, N. &. Simons, R.K. (2008). The spatial epidemiology of trauma: the potential of geographic information science to organize data and reveal patterns of injury and services. Canadian *Journal of Surgery*, 51, 389 (2008).

Shalowitz, D., & Wendler, D. (2006). Informed consent for research and authorization under the Health Insurance Portability and Accountability Act: An integrated approach. *Annals of Internal Medicine*, 144, 685-688.

VanWey, L.K., Rindfuss, R.R., Gutmann, M.P., Entwisle, B., & Balk, D.L. (2005). *Confidentiality and Spatially Explicit Data: Concerns and Challenges*. Proceedings of the National Academy of Sciences of the United States of America 15337, 102, 15337- at 15341.